The

PIONEER RESEARCH

Journal

An International Collection of
Undergraduate-Level Research

Volume 9
2022

Pioneer®
academics

Publisher
Hong Kong Digital Publishing Ltd.
1501, Grand Millennium Plaza, 181 Queen's Road, Hong Kong

ISBN 978-988-77066-3-2

Pioneer Academics, PBC
Email: info@pioneeracademics.com
Website: www.pioneeracademics.com

The Pioneer Research Journal is published annually by Pioneer Academics, PBC. The 2022 issue of *The Pioneer Research Journal* is Volume 9.

Contents

Foreword

Throughout the pages of Volume 9 of the *Pioneer Research Journal*, 27 outstanding high school students demonstrate their commitment to intellectual and social engagement with the world around them. From major environmental problems to the history of the AIDS epidemic, these papers engage with issues of worldly consequence and with ongoing debates of scholarly importance. As in recent editions, the COVID pandemic remains a prominently featured topic in this volume, while being examined from the perspective of an ever-increasing array of disciplines, ranging from biology to economics to gender studies.

The work featured in the pages that follow represents the culmination of the authors' participation in the Pioneer Research Program. In 2022, 1,464 young scholars from 51 countries and regions conducted research through Pioneer Academics, selected from an international pool of 4,775 applicants. Those admitted to the program participated in a faculty-led, international cohort before working individually with leading U.S. professors to conduct original undergraduate-level research in their area of interest. After a rigorous nomination and double-blind review process, 27 papers were selected for publication in this edition of the *Pioneer Research Journal.* The authors represented herein are from schools in Cambodia, Canada, China, South Korea, Turkey, the United States, and Vietnam.

Oberlin College's collaboration with Pioneer Academics to ensure intellectual rigor and to uphold the highest standards among faculty and scholars alike is driven by a shared commitment to excellence and access. By conducting the program entirely online, Pioneer has torn down barriers and made undergraduate-level education available to promising young scholars in virtually every corner of the world. To ensure this opportunity is available to as many students as possible, Pioneer provided US $1.46 million in need-based scholarships, again meeting 100% of demonstrated financial need where need could be assessed and fulfilling their mission to remove obstacles to educational access for the most deserving underserved students.

The work contained in this journal clearly demonstrates what can be achieved when expert guidance and a rigorous academic approach intersect with the deep intellectual curiosity and passion possessed by the young authors whose papers are featured in this volume.

It is our goal to share this work widely, and so the *Pioneer Research Journal* is available in print and online at www.pioneeracademics.com and is distributed to select colleges, universities, and libraries worldwide.

I am so pleased to share it with you and hope you find it to be inspiring and enlightening.

David G. Kamitsuka, Ph.D.
Dean of the College of Arts and Sciences
Oberlin College & Conservatory
Editor, The Pioneer Research Journal, Volume 9

Selection Process

To be nominated for publication in *The Pioneer Research Journal*, Pioneer scholars first had to earn a letter grade of A- or higher from their professor mentor. Each professor was invited to nominate one or two papers (depending on the number of scholars they mentored) that met the A/A- grading threshold and that, in their estimation, represented the highest caliber of scholarship from among their mentees.

Following nomination, every paper underwent a holistic review, which included an assessment of each scholar's academic environment provided by their school as well as a double-blind evaluation by a member of our 50-member committee of contributing readers, each of whom is a professor with expertise in the academic discipline of the papers they reviewed.

Contributing Readers scored the papers on a scale of 0 – 6 for four criteria: Engagement with Scholarship, Evidence & Analysis, Writing & Organization, and Scholarly Contribution. These four scores were then tallied, with a possible maximum score of 24. Papers with sub-scores of 1 in any category were disqualified from publication.

Once the above thresholds were met, the highest-scoring papers were provisionally selected. Upon notification, authors of provisionally selected papers received instructions from their paper's reviewer specifying the revisions required for meeting the standards of an undergraduate research journal. Every paper then underwent one more rigorous review by Pioneer's Writing Center tutors for editorial concerns and the authors were asked to make final revisions based on that review before their papers were finally published.

When all criteria were applied, only 1.9% of the papers generated in the 2022 Spring-through-Summer and Summer-Only terms were selected for publication in Volume 9 of *The Pioneer Research Journal.*

Adaptive Reuse of Temporary Structures for a Humanitarian Purpose: A Case Study of the Jarahieh School

Kate Kim

Author Background: *Kate Kim grew up in South Korea and currently attends the Peddie School in Hightstown, New Jersey in the United States. Her Pioneer research concentration was in the field of architecture and titled "Sustainable Approaches to Architecture Heritage."*

Abstract

Participatory design, sustainable design, and community-led construction are the three approaches utilized in designing and implementing the Jarahieh School, a project spearheaded by CatalyicAction, a U.K.-based not-for-profit design studio organization. The project commenced in December 2015 and was completed in October 2016 for a humanitarian cause in the Syrian refugee settlement of Jarahieh located in Al Marj, West Bekaa, Lebanon. The participatory design approach is based on the principle of social inclusion and collaboration with the beneficiaries and other stakeholders during the project's design and conception of ideas. The sustainable design approach rests on recycling and reusing temporary structures which would otherwise be discarded or made obsolete and makes use of them as construction materials while adapting to the local context. The community-led construction approach seeks to achieve the goal of installing a sense of ownership and empowerment on the part of the beneficiaries as well as the local community by engaging them directly in the construction and implementation phases of the project. This paper explores the Jarahieh School as an innovative case of sustainable adaptive reuse and argues that the three approaches serve as an ideal framework for how sustainable adaptive reuse can potentially be applied inventively and effectively in projects involving the use of temporary structures to help address the world's ongoing humanitarian crises.

1. Introduction

Broadly, adaptive reuse is the process of extending the useful life of old, historic, or obsolete buildings.[1] Most of the literature and research studies on adaptive reuse have focused on models or applications of adaptive reuse in the context of urban regeneration or commercial redevelopment.[2]

There is a lack of scholarship and research delving into how adaptive reuse can potentially be applied to projects aimed at addressing a humanitarian crisis. In particular, the ongoing Syrian refugee crisis is considered one of the most serious humanitarian crises in modern history.[3] Although makeshift tents provide temporary housing for persons living in refugee settlements, this is only a short-form solution. On a mid- or long-term basis, a different, innovative solution may be required to respond to the complex circumstances brought about by forced displacement, especially to meet the wellbeing and social needs of refugees and forcibly displaced persons in humanitarian crises.

Adaptive reuse of temporary structures has the potential to improve the welfare of the people suffering from a humanitarian crisis. This paper explores whether and to what extent adaptive reuse of temporary structures can be used as a viable strategy for a humanitarian purpose, through a case study of the Jarahieh School, a project that entailed the construction of a school in 2016 in the Syrian refugee settlement of Jarahieh located Al Marj, in the western area of Bekaa region, Lebanon (see Figure 1).

The overall design and implementation of the Jarahieh School project were spearheaded by CatalyticAction, a U.K.-based not-for-profit design studio organization. Through the organization's collaboration with various stakeholders, the project entailed transporting temporary exhibition structures that were displayed at the Milan Expo 2015 from Italy to the Jarahieh refugee settlement in Lebanon and recycling and reusing the structures' materials to construct a multi-purpose school at the refugee settlement site (see Figure 2).

Considering the project's history, relevant economic, social, and environmental considerations, and the processes through which the project was conceived, designed, and implemented, the Jarahieh School project is an innovative case of adaptive reuse. But more importantly, the project is unique in that it applied participatory design, sustainable design, and community-engaged construction. This paper highlights the three approaches adopted in the Jarahieh School project as an ideal framework for designing and implementing future

---

[1] Shieda Shahi, Mansour Esnaashary Esfahani, Chris Bachmann, and Carl Haas, "A Definition Framework for Building Adaption Projects," *Sustainable Cities and Society* 63 (2020): 9.

[2] Peter A. Bullen and Peter E.D. Love, "Adaptive Reuse of Heritage Buildings," *Structural Survey* 29, no. 5 (2011): 411–421; Craig Langston, Francis K.W. Wong, Eddie C.M. Hui, and Li-Yin Shen. "Strategic Assessment of Building Adaptive Reuse Opportunities in Hong Kong," *Building and Environment* 43, no. 10 (2008): 1709–18; Peter J. Larkham, "Rebuilding the Industrial Town: Wartime Wolverhampton," *Urban History* 29, no. 3 (2002): 388–409.

[3] Daniela V. Dimitrova, Emel Ozdora-Aksak, and Colleen Connolly-Ahern, "On the Border of the Syrian Refugee Crisis: Views from Two Different Cultural Perspectives," *American Behavioral Scientist* 62, no. 4 (February 2018): 532–46.

adaptive reuse projects involving temporary structures in a humanitarian crisis context.

Figure 1. *Map of Lebanon showing the West Bekaa region, the Al-Marj town within West Bekaa and the location of the Jarahieh School (source: United Nations for High Commissioner for Refugees (UNHCR), September 2016 and Google Earth).*

Figure 2. *A visual representation of the transportation route of the pavilion structure from its original location in Milan, Italy to its current location in Al-Marj, Lebanon (source: public domain and modified by author). The pavilion structure was displayed at the Milan Expo of 2015 and was later donated to a charity, which was used to construct the Jarahieh School located in the Syrian refugee settlement in Al-Marj, Lebanon.*

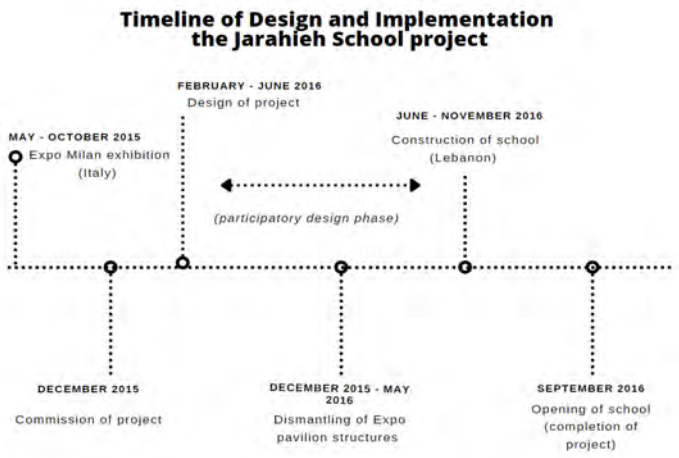


Figure 3. *A timeline of design and implementation milestones of the project (source: figure by author).*

2. Adaptive Reuse – Overview

2.1 What Constitutes Adaptive Reuse?

According to one study authored by Shelda Shahi et al. in 2020 that analyzed various literature and research studies on different types and forms of building adaption projects (the "Shahi Study"), the term "adaptive reuse" is defined as "the process of reusing an obsolete and derelict building by changing its function and maximizing the reuse and retention of existing materials and structures."[4] The Shahi Study is noteworthy in that it analyzed "over 1,600 papers published from 2011 to 2020 involving selected terminologies, including retrofitting, renovation, rehabilitation, refurbishment, material reuse, building conversion, and adaptive reuse."[5] The Shahi Study's authors note that terminologies of refurbishment, retrofitting, rehabilitation, renovation, restoration, modernization, conversion, adaptive reuse, material reuse, conservation, and preservation are frequently used interchangeably due to coinciding or interrelating scope and deficient clearness for their appropriate or suitable usage, and the study's objective is to "develop a definition framework that avoids costly confusion by enabling clear and consistent use of building adaptation terms based on the characteristics and scope of each project."[6]

According to the Shahi Study, at the highest level, it distinguishes building adaptation projects between refurbishment and adaptive reuse. Within the adaptive reuse category, the first sub-category of "conversion" describes "changing the function of a building or some parts of the building," and the second sub-category of "material reuse" describes "recovering and reusing existing

[4] Shahi, *Definition Framework for Building Adaptation Projects*, 2.
[5] Ibid.
[6] Ibid.

materials of a building" (see Figure 4).

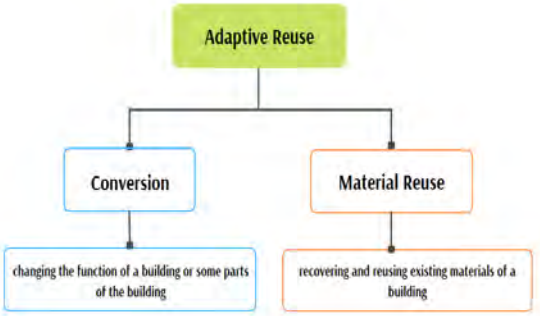


Figure 4. *Graphic illustrations of two different variants of adaptive reuse: (i) changing the function of a building or some parts of the building (as conversion) and (ii) recovering and reusing existing materials of a building (as material reuse) (source: figure by author based on illustrations provided in Shahi, 2020).*

Hence, according to the definitional framework in the Shahi Study, either a project entailing a "conversion" or "material reuse" would qualify as an adaptive reuse project. As discussed in Section 3, the Jarahieh School project constitutes a case of adaptive reuse, whether based on applying the definition of "conversion" or "material reuse" (or both) proposed in the Shahi Study (see Figure 5).

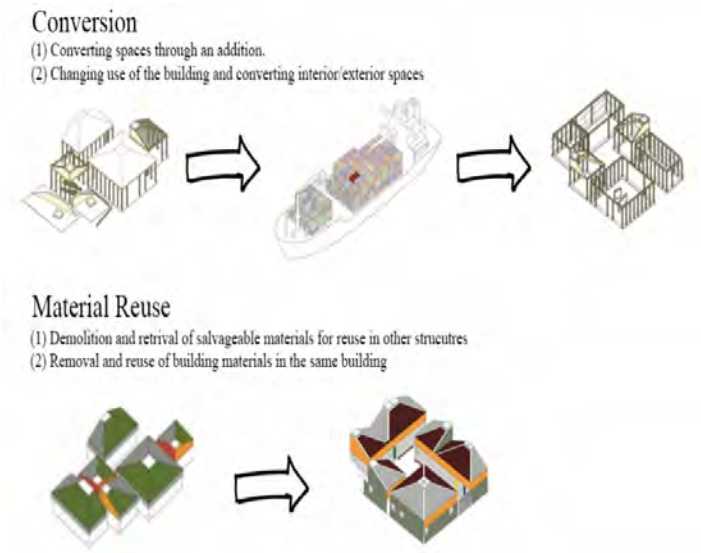


Figure 5. *Visual diagrams showcasing examples of "conversion" and "material reuse" in the case of the Jarahieh School (source: modified figure by author based on illustrations provided in Dabaj and text descriptions provided in Shahi).*

2.2 General Trends of Adaptive Reuse

In academia, according to one recent bibliometric analysis study that examined more than 200 journal articles published from 2006 to 2021 regarding global research developments in adaptive reuse, a great majority of them were published in the past four years (2019 through 2021), engineering and environmental science were the top two subject matters researched, and the issues and topics studied in the articles pertained to building and construction adaptive reuse in the areas of (i) material reuse, (ii) assessment of life cycle, (iii) economic assessment, (iv) multi-criteria decision making, (v) regulatory policies, and (iv) stakeholders' analysis.[7]

In industry, on the other hand, emerging themes and trends in adaptive reuse have focused on, among others, re-purposing older buildings for mixed uses, restoration and reuse of heritage buildings, and projects in metropolitan areas for urban renewal and regeneration purposes.[8]

2.3 Examples of Adaptive Reuse Projects

Some common examples of sites subject to adaptive reuse include converting culturally, historically, or architecturally significant buildings into community centers or mixed-use creative venues that would otherwise be left to decay or demolished.

In the United States, the first major adaptive reuse project was Ghirardelli Square in San Francisco, which was completed in 1964.[9] This once neglected one-block complex containing a former chocolate factory was transformed into a shopping and tourist destination, which served as a viable adaptive reuse model for other cities in the U.S. and has become one of the top tourist spots in the city of San Francisco.[10]

Outside of the United States, another recent example of adaptive reuse is the Senate of Canada Building, located in Ottawa, Canada. In this 2019 adaptive reuse project, a train station was converted into a government building, which also entailed the reuse of existing building structure materials.[11]

Adaptive reuse projects need not always be momentous or on a large scale in terms of the built environment. Some examples of modest adaptive reuse projects include (i) Café Restaurant Amsterdam in Amsterdam, the Netherlands, which is operated inside a former water-processing plant built in the 1800s where

---

[7] Oluwatobi Owojori, Chioma Okoro and Nicholas Chileshe, "Current Status and Emerging Trends on the Adaptive Reuse of Buildings: A Bibliometric Analysis," *Sustainability* 13 (2021): 1-18.

[8] Steve Knaub, "The Potential of Adaptive Reuse," Bill Gladstone Group. NAI Commercial-Industrial Realty Co., August 14, 2020.

[9] "Ghirardelli Square," The Landscape Architecture of Lawrence Halprin (The Cultural Landscape Foundation).

[10] Ibid.

[11] Akiva Blander, "Diamond Schmitt Architects Adapts a Historic Train Station for the Canadian Senate," *Metropolis*, March 29, 2019.

the structure remains left mostly intact,[12] and (ii) XY Yunlu Hotel, located in Guangxi, China, which involved converting five dilapidated farmhouses to make guest rooms for a new high-end boutique hotel.[13]

Ghirardelli Square, San Francisco, USA Senate of Canada, Ottawa, Canada

Figure 6. *Photographs of Ghirardelli Square and the Senate of Canada (source: the American Society of Landscape[14] Architects and Saffron Blaze,[15] respectively).*

Café Restaurant Amsterdam, Amsterdam, Netherlands XY Yunlu Hotel, Guangxi, China

Figure 7. *Photographs of Cafe Restaurant Amsterdam and XY Yunlu Hotel (source: The Design Gesture [16] and dezeen,[17] respectively).*

3. The Jarahieh School: a Case of Adaptive Reuse Project

Based on the definition of adaptive use proposed in the Shahi Study, the Jarahieh school project qualifies as an adaptive use project. It falls into both the "conversion" sub-category and the "material reuse" sub-category. While the

[12] Florian Heilmeyer, "6 Projects That Made the Netherlands a World Capital of Adaptive Reuse," *Metropolis*, February 17, 2021.
[13] Lizzie Crook, "Atelier Liu Yuyang Reuses Old Farmhouses to Create Boutique Hotel in Rural China," *dezeen*, October 15, 2019.
[14] "Ghirardelli Square."
[15] Helen Forsey, "Excerpt: Envisaging a People's Senate," Canadian Centre for Policy Alternatives, April 1, 2015.
[16] Eeti Goel, "Adaptive Reuse Architecture: Breathing New Life in Structures," *The Design Gesture*.
[17] "XY Yunlu Hotel by Atelier Liu Yuyang: Dezeen Awards (Shortlist for Hospitality Building of the Year)," dezeen (2019).

Jarahieh School project could arguably be viewed as a mere instance of recycling waste or by-products, there are aspects of building conversion and material reuse associated with the project, which makes it an inventive case of adaptive reuse.

3.1 The "Conversion" Aspect

The project was unique in that it did not entail any use of the building or structure already existing at the site. The original structure came from Italy and was a pavilion in the form of a temporary exhibition, of which construction was commissioned by the Italian branch of Save the Children, an international non-governmental organization (NGO), for the 2015 International Expo.[18] Although the pavilion was constructed for the exhibition purpose and considered obsolete after the exhibition, the association intended that the structure be reused afterward for a charitable purpose. According to Save the Children, the architectural reuse of the expo pavilion was a commitment made as part of its philanthropic mission and CatalyticAction, the architect of the site, collaborated with it to design and accomplish the pavilion's "afterlife."[19]

The pavilion's modular design and its overall configuration of the structure were substantially maintained, but specifically converted for new use (i.e., school) and adapted to the local context (see Figure 8). The original Expo pavilion did not have any partition walls and contained a set of columns that were free-standing to support six modular structures. The adapted structure of the school is based on the same six modular structures but enclosed with robust walls and reorganized in a circular fashion. The six modular structures were reconfigured to meet the needs of the school and the beneficiaries of the settlement, where the reconfigured structures provided a space that served as an outdoor courtyard for the education and physical activities of school children.[20]

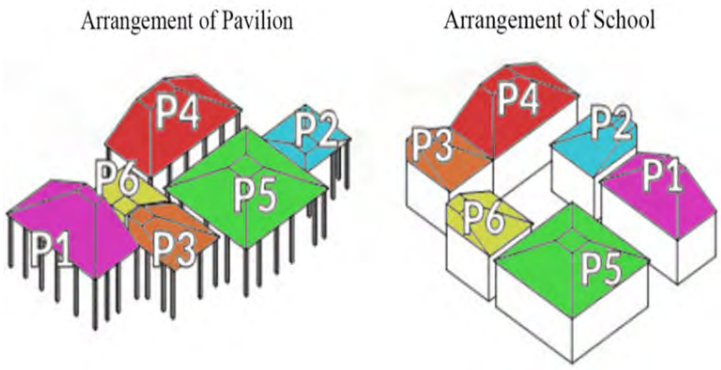


Figure 8. *Color-coded diagrams of the modular arrangement of the original pavilion structure set opposite the Jarahieh School (source: modified figure by author based on illustrations provided in LafargeHolcim). There are a total of six modules in the original pavilion,*

[18] Berlanda, Toma, "2019 On Site Review Report: Jarahieh School," 1.
[19] "Repurposed Exposition Pavilion," Piggy Backing Practices (a virtual symposium) of University of Arkansas, School of Architecture + Design, 1.
[20] Ibid.

which were all recycled and reused to construct the Jarahieh School. In the original pavilion structure, the modules were arranged in a clustered linear format, while for the school, the modules surround a central courtyard.

There are other notable factors supporting the "conversion" aspect of the project. In particular, the original pavilion structure was built for an exhibition purpose but was subsequently transformed to be earthquake resistant. Specifically, by working with a third-party design firm, steel seismic anchor plates were newly fabricated, and these were connected to the foundations of the vertical structure. The design firm made "specific designs for steel plate shear connectors and sole plates, together with lateral bracing, were provided"[21] and these apparatuses were manufactured and installed at the school site for seismic load resistance. This made the school the only earthquake-resistant structure in the settlement. Further, the original pavilion structure with wooden frames was substantially retained, which ended up dictating the geometry of the six modular structures serving as the school's classrooms, but it was turned into an indoor space complex that was laid around a courtyard, which created a village within the village settlement site.

3.2 The "Material Reuse" Aspect

The pavilion structure consisted of timber, wood panels, and iron sheets, and these materials were all salvaged and reused for the construction of the school. They were dismantled, packaged, and shipped from Italy to Lebanon to be recycled for new, sustainable, and better utilization. Once arrived at the site, the materials were assembled and supplemented with local materials for insulation, which required a set of local skills and knowledge of the materials as the basis of overall design and construction. By reusing the materials of the original building, the building was "reclaimed" to serve an entirely different purpose as a multi-purpose school (which also functions as a community center). The material reuse aspect of the project is inventive in that it is "an example of waste-stream piggybacking that crosses continents and cultures to capture the architectural by-products from an international exposition."[22]

4. Background and History of the Jarahieh School Project

4.1 The Syrian Humanitarian Crisis in Lebanon

In 2016 (when the project was completed), Lebanon hosted 1,800,000 refugees from Syria who fled from their homes due to war, which represented close to a third of the total population of the country (see Figure 9). Children made up more than 50% of these Syrian refugees and they were not able to receive education

[21] Berlanda, Toma, "2019 On Site Review Report: Jarahieh School," 5.
[22] Ibid.

from the Lebanese government.[23] In 2016, according to Human Rights Watch, it was reported that 250,000 Syrian refugee children in Lebanon were not receiving any school education."[24]

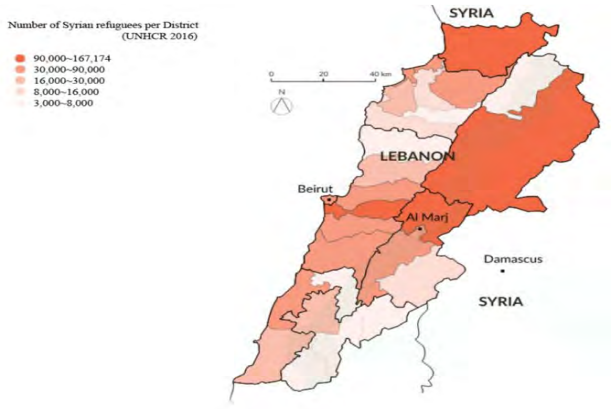


Figure 9. *A color-coded map of Lebanon showing the number of Syrian refugees per district (source: modified figure by author based on an illustration provided in "Pavilion Re-Claimed in Lebanon").*

The Jarahieh refugee settlement was established by United National High Commissioner for Refugees (UNHCR) in 2013. Located on the eastern part of the city of Al Marj in the West Bekka region, the settlement then comprised about 300 tents with a surface size of roughly 30,000 square meters.[25]

In 2016 (when the project was completed), the refugees lived in harsh conditions. Makeshift tents were built and assembled in different materials, comprising mainly wood, tarpaulin (heavy-duty waterproof cloth, originally of tarred canvas), and metal sheets.[26]

4.2 The Original School

For young children, there were no safe places to play in their surroundings. The Syrian refugee crisis has negatively impacted their mental well-being.[27] As a response to the lack of education and educational opportunities, Sawa for Development Aid (a non-profit organization) and Jusoor (a non-profit organization) originally set up a school in the refugee settlement of Jarahieh, established by the UNCHR. Built using wooden frames and draped with fabric, the previous school was constructed as a tent in a similar manner to temporary

[23] "CatalyticAction: Jarahieh School for Syrian Refugee Children in Lebanon," Floornature Architecture & Surfaces, June 11, 2019.
[24] Ibid.
[25] Berlanda, "On Site Review Report," 2.
[26] Ibid.
[27] Beatrice De Carli, Celia Macedo, and Lucia Caistor-Arendar, "CatalyticAction: Interview of Joana Dabaj," *Pedagogies of Inclusion Vol.1: A review of Spatial Design Education in Europe (2019).*

housing shelters provided by the UNHCR.

The school provided education to around 320 children ranging from ages five to fourteen each year. However, problems regarding the former school were various and often disruptive to the well-being and safety of the beneficiaries. For example, it "was structurally inadequate in that it could not, for example, protect children against the mud that accumulated on the land in rainy weather, it was poorly lit, and it gave an impression of temporariness."[28] The tented school was subject to poor lighting, issues regarding local climate/temperature, sound levels, and a lack of educational recreational facilities (see Figure 10).[29] In addition to these shortcomings, the school was required to undergo unnecessary annual costs due to problems relating to the maintenance of the school.

Figure 10. *A photograph of children of the Jarahieh site taking classes in their previous dimly lit school, constructed from a large tent structure (source: Berlanda, 2019).*

4.3 The Redevelopment

Beyond addressing issues relating to poor lighting, temperature and sound levels, and a lack of any recreational space, the process of conceiving ideas and designs for the redevelopment of the school was driven by the need to improve and enhance the educational environment and the quality of life for these children and their families, such that the school should not only be a place to learn skills, but also a psychological and emotional safe haven, and also help to cultivate a community of cohesion within the Syrian refugee settlement itself and within the Lebanese society more broadly (see Figure 11). The main objective was to open "the door for Syrian refugee children to a comprehensive remedial education,

[28] "CatalyticAction: Jarahieh School for Syrian Refugee Children in Lebanon."
[29] "Temporary Expo Structure Is Repurposed as School for Syrian Refugees in Lebanon," *REVITALIZATION: The Journal of Urban, Rural & Environmental Resilience, 60,* September 17, 2017.

which will then enable them to successfully enroll in public schools in Lebanon."[30]

Figure 11. *Two bird's eye view images of the proposed site of the Jarahieh school (source: modified figure by author based on images provided in Google Earth and Berlanda, 2019). Both images are accompanied by detailed photographs of the specific parts of the site that are expressed through the arrows. The image on the left shows the location of the Jarahieh School in 2014 before any construction took place. The image on the right shows the Jarahieh School in 2017, a year after construction was completed. As shown through the comparison of the two images, the Jarahieh School replaced the previous tented school as shown in highlighted regions portrayed in each of the images.*

Moreover, due to a dearth of basic needs and opportunities for livelihood for residents in the Jarahieh refugee settlement, the redevelopment of the school focused on not only serving as educational facilities for children, but also as a center of community activities and the only safe and secure facility in the event of snowstorm or earthquake in the settlement.[31] As part of the overall design, the school was intended to not just serve as a school for children; after 4 pm it is a school for adults, and on weekends it functions as a public cinema and a site for aid distribution. The space formed between the six modular buildings on the school site was intended to create a public area for all residents of the refugee settlement, and during a potential natural disaster such as a flood or snowstorm, the buildings were intended to double as a community shelter."[32]

[30] Ibid.
[31] Berlanda, "On Site Review Report," 9.
[32] "Jarahieh School, Al-Marj, Lebanon," *Archnet*.

5. The Architect: CatalyticAction

CatalyticAction is a U.K.-based non-profit design studio organization "that works to empower communities through strategic and innovative spatial interventions," focusing on refugee and displaced populations in the Middle East since 2015. Established in 2010, it participates in designing projects that "catalyze" to bring about positive change in society through architectural design.[33]

When working on projects, CatalyticAction's priorities are creating a sense of belonging. According to CatalyticAction, they focus primarily on "solidity, structural simplicity, rapid construction of the building, and the possibility of obtaining materials among the community that would be using the [building]."[34]

According to CatalyticAction, participatory design is one of its key working philosophies. Over the years, CatalyticAction has developed a set of participatory tools that work most often in most places, but they are open to new ideas from their partners, from the community, and anyone else who takes part in the collective design process of each project.

In terms of engaging with the community, they not only collaborate with various stakeholders, but also work closely with local partners, who have in-depth knowledge of the place and the local community. According to CatalyticAction, community ownership of a project is a key factor and this is achieved as a result of going through all phases of idea conception, design, and implementation of the project together with the community, including women, children, and youth, alongside experienced builders, who all participate and contribute to the project's process and progress.

For the Jarahieh School project, CatalyticAction collaborated with various stakeholders: (i) Jusoor NGO running the educational program at the school site, (ii) Save the Children donating the pavilion structure, (iii) Sawa for Development & Aid for local support, (iv) ARUP providing engineering pro bono consultancy, (v) Argot ou La Maison Mobile providing technical assistance on the pavilion structure design, (vi) the community members of the Jarahieh refugee settlement, and (vii) the support of local and international volunteers."[35]

Examples of other successfully completed projects of CatalyticAction are (see Figure 12):

- School Playgrounds Rehabilitation with Right to Play, Lebanon (2019~2020);[36]
- Safe Parks in Kaskas and Basta with Himaya, Beirut, Lebanon (2021); and
- Karantina Park Rehabilitation with Terre des Hommes, Beirut, Lebanon (2019~2020).

[33] "CatalyticAction: Jarahieh School for Syrian Refugee Children in Lebanon."
[34] Ibid.
[35] Berlanda, "On Site Review Report," 3.
[36] "CatalyticAction: Jarahieh School for Syrian Refugee Children in Lebanon."

School Playgrounds Rehabilitation with Right to Play, Lebanon (2019~2020)

Safe Parks in Kaskas and Basta with Himaya, Beirut, Lebanon (2021)

Karantina Park rehabilitation with Terre des Hommes, Beirut, Lebanon (2019~2020)

Figure 12. *An image collage of the three recent, successful projects completed by CatalyticAction (source: CatalyticAction).*

6. Participatory Design Approach to the Jarahieh School Project

One of the three approaches utilized in the Jarahieh School project is participatory design (see Figure 13). Participatory design is an approach to design that actively involves all relevant stakeholders in the design process to help ensure the result meets their needs and is usable.[37]

[37] "Jarahieh School," CatalyticAction.

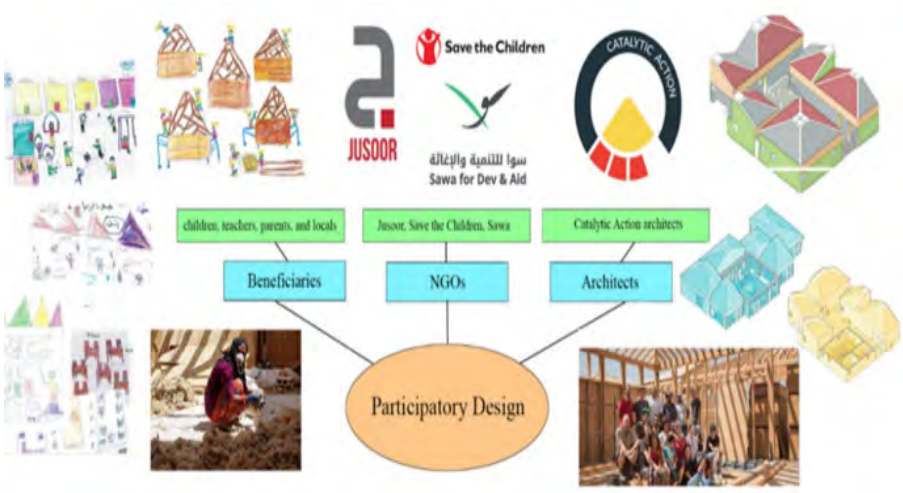


Figure 13. *Diagrams and graphics depicting the various stakeholders in the process of participatory design implemented in CatalyticAction's planning and construction of the Jarahieh School (source: modified figure by author based on illustrations and information provided in "Pavilion Re-Claimed in Lebanon"). Participatory design, one of the three approaches advocated by CatalyticAction, endeavors to produce an ideal product through the combined effort of the beneficiaries, NGOs, and architects. The beneficiaries of this project are the locals; the NGOs are Jusoor, Sawa for Development and Aid, and Save the Children; and the architects are CatalyticAction.*

By engaging in participatory design practices, the main aspects of the goal of the project were to: create a design to lower the school's running costs; extend the range of activities the school can provide; keep warm in winter and cool through the summer; allow light into the classrooms, and prevent noise from traveling.[38] Also, due to a diverse range of Syrian ethnic groups living within the Jarahieh informal settlement, the participatory design presented a unique opportunity for the beneficiaries to work together on what would become a "shared piece of infrastructure."[39]

The participatory design activities carried out focused on understanding the children's, teachers', and parents' needs and desires, which helped to increase their sense of ownership over the project. These activities comprised workshops, focus group meetings, and interviews to engage with children, NGOs, municipality members, teachers, and parents. For example, through a series of workshops, both prior to the arrival of the building structure components from overseas, and once these materials were on site, input was sought on the different elements of the school design. This allowed for the needs and desires of the beneficiaries and stakeholders to be considered to facilitate the design, resulting in the modification and customization of the physical configurations and location of the school structures and the classroom's arrangement.[40]

[38] Ibid.
[39] Ibid.
[40] Ibid.

As a result of applying participatory design practices, after the completion of the school construction, several technical improvements in the operation of the school were made possible, as follows: (i) drinking water in storage and distribution tanks, (ii) spaces free of insects and rodents, (iii) additional storage spaces, (iv) bathrooms and sinks that are connected to the sewerage system, and (v) classrooms that are well-lit and thermally comfortable.

In particular, prior to the school's redevelopment, one of the significant issues that the beneficiaries had to endure was poor lighting in the old school building. The new school was designed in such a way as to provide natural skylight from the ceiling to provide bright and naturally lit classrooms. Exceptionally, adequate natural light was not possible during the winter months and electrical fluorescent tubes are instead used for lighting inside the classrooms.

Also, in order to provide a "playful, stimulating environment for hundreds of young people who are forced to endure life in the harsh conditions of the refugee settlement," the school was designed to provide dedicated recreation spaces[41] and the outside classroom walls were covered with educational and decorative murals made by the inhabitants themselves, avoiding the "anonymous look characteristic of the streets of tent cities and offering a symbol of community pride."[42]

From a design standpoint, the use of basic and recognizable materials, such as wood, helped to create a bond between the beneficiaries and their families, and the school. Instead of constructing a single big building with internal subdivisions, the six modular structures are made up of separate classrooms arranged around a courtyard. They were built on the same scale as the homes. To modify the classrooms and give a unique identity of its own for each the classrooms, it is made in such a way that it is not the same and different from others, with roofs based on pyramid shapes positioned at varying heights and angles."[43]

7. Sustainable Design Approach to the Jarahieh School Project

The second approach utilized in the Jarahieh school project was sustainable design. Sustainable design seeks to reduce negative impacts on the environment, and the health and comfort of building occupants, thereby improving building performance.

One main aspect of the sustainable design approach utilized in the project rested on recycling and reusing the pavilion structures which would otherwise be discarded or made obsolete. They included wood timber, wood panels, and iron sheets and were reused to construct benches in the school's courtyard, climbing walls and bathroom walls at the school site, and protective rain screens and roofing sheets, respectively.

In addition to making use of the pavilion structures as construction materials, the other main aspect of the sustainable design was the strategy of

[41] "Temporary Expo Structure Is Repurposed as School for Syrian Refugees in Lebanon."
[42] Ibid.
[43] Ibid.

adapting to the local context in terms of locally sourcing the construction that was cost-effective and environmentally sustainable. To illustrate, construction materials locally sourced were as follows:[44]

Table 1. *A summary table of various sustainable materials utilized in the construction of the Jarahieh School (source: table by author based on information provided in "Pavilion Re-Claimed in Lebanon" and Berlanda, 2019). The first column indicates the materials used, the second column describes the function of the materials, and the third column states the source of the materials.*

Material	Function	Procurement source
Green mesh	- wool insulation to breathe and provided structure - local community uses it for the protection of vegetation	- procured from the local community
Grains bag fabric	- bulk sold material that is economical and sustainable	- procured from the local community
OSB (Oriented Strand Board) panels	- building wall materials	- supplied by the UNHCR for the refugees
Local iron sheets	- building wall materials	- procured from local farms
Sheep wool	- infill materials (insulation for wall and ceiling, and membranes that are damp-proof) - protect children from the summer heat and winter cold while providing noise insulation - natural wool acts as an excellent sound buffer between classrooms and is an extremely efficient thermal barrier - completely sustainable and requires far less energy to produce than the equivalent human-made product	- procured from local farms (sheep wool is considered a waste product)

A graphical illustration of the innovative uses of suitable technologies, labor, and materials based on environment and climate conditions is shown below (See Figure 14).

[44] "Pavilion Re-Claimed in Lebanon," Holcim Foundation.

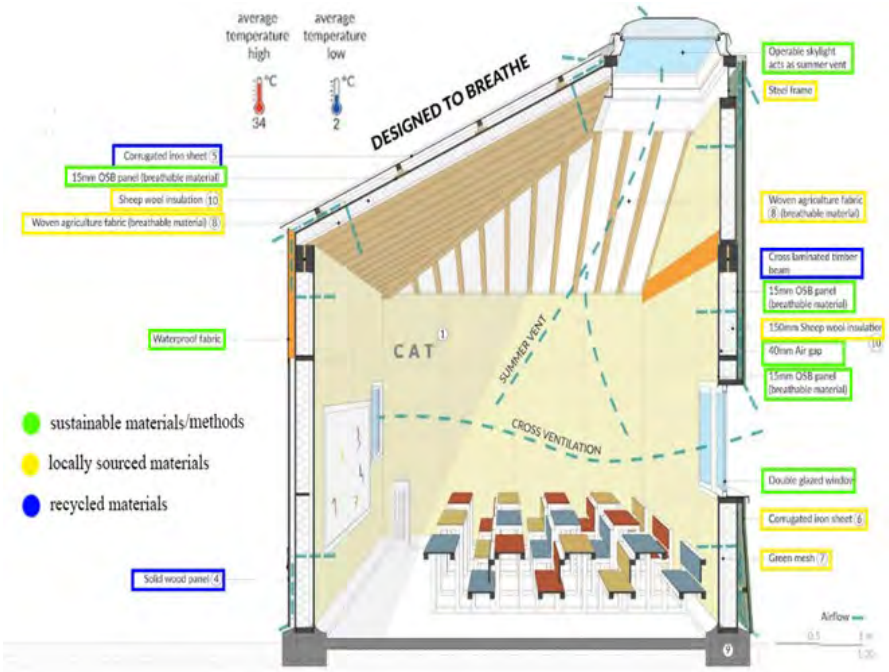


Figure 14. *A modified diagram of the original diagram of materials used in the construction of the Jarahieh School (source: modified by author based on illustrations provided in "Pavilion Re-Claimed in Lebanon"). The diagram is color coded to display what materials were locally sourced, which ones were recycled, and various sustainable materials as well as methods used to promote the greatest possible energy efficiency and environmental friendliness. The diagram is also accompanied by a key that shows the meanings of three colors and a graphic of the average temperature highs and lows at the site of the school.*

In particular, the walls and roofs of the school are insulated with locally sourced sheep's wool to help the building breathe air and provide humidity control. Acting as thermal insulation and a sound barrier, the sheep wool functions to protect children from varying temperatures and noise traveling between classrooms. The use of local sheep wool also entailed the process of working with local women and farmers to gain their know-how and expertise when hiring them. For example, with specific knowledge of how to clean, dry, and prepare the wool for use, women contributed greatly to the construction process, a process which is typically dominated by men.[45] Through methods previously stated, local women were encouraged to participate in the project and were empowered in the process of doing so.

[45] Ibid.

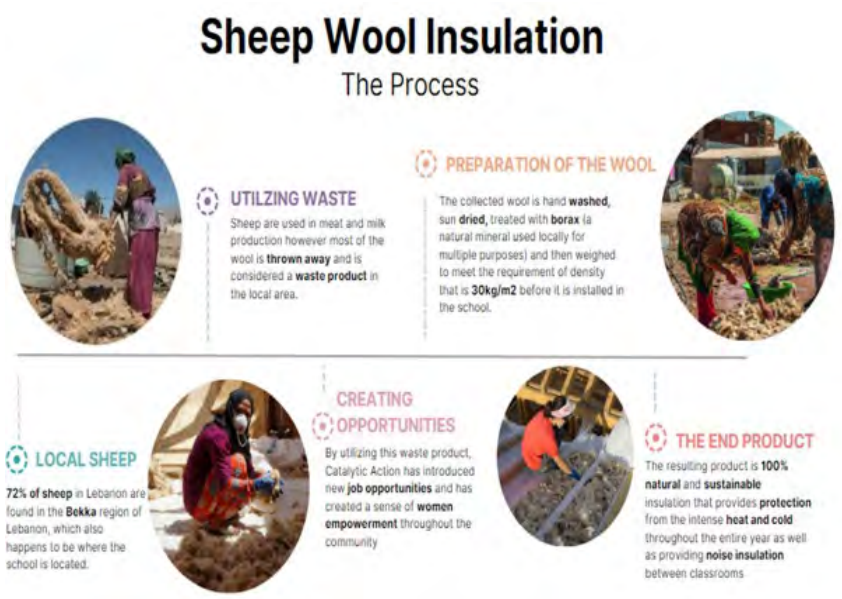


Figure 15. *This figure illustrates the process of creating the sheep wool insulation that was used in the construction of the Jarahieh School (source: modified by author based on illustrations provided in "Pavilion Re-Claimed in Lebanon"). It outlines the most important steps and shows various context information as well as the significance of the process.*

8. Community-Led Construction Approach to the Jarahieh School Project

Community-led construction was the third approach utilized in the Jarahieh School project. According to CatalyticAction, it is a process that seeks to achieve the goal of installing a sense of ownership and empowerment on the part of the beneficiaries as well as the local community at large by engaging them directly in the project's construction.[46]

There are several tangible benefits to the community-led construction approach. One obvious benefit is the cost saving realized by not having to hire professional contractors. For the project's construction, there was no main contractor hired, and the construction team of the project was comprised of mostly Syrian refugees residing in the settlement, craftsperson workers and local artisans from the nearby community, and volunteers from various NGOs.[47]

The second tangible benefit is the opportunity to directly transfer skills and knowledge to local laborers from the Jarahieh community who participated in the construction activities, equipping both women and men with the resources to continue to work for the community beyond this single project.[48] This social inclusion practice of sharing and transferring knowledge aimed to "bridges in a

[46] "Jarahieh School," CatalyticAction.
[47] Berlanda, "On Site Review Report," 6.
[48] "Jarahieh School," CatalyticAction.

community that is made up of a wide range of Syrian ethnic groups."[49] For example, in the project's construction phase, 26 local businesses were involved, 53 local workers were employed (of whom 13 were women), and 27 local youth were trained.[50]

Thirdly, community-led construction has the benefit of assisting with post-completion employment and establishing connections and relationships with the members of the local community. After the school was constructed, 12 teachers were employed, and four businesses were engaged as all local materials were procured from the nearby communities.

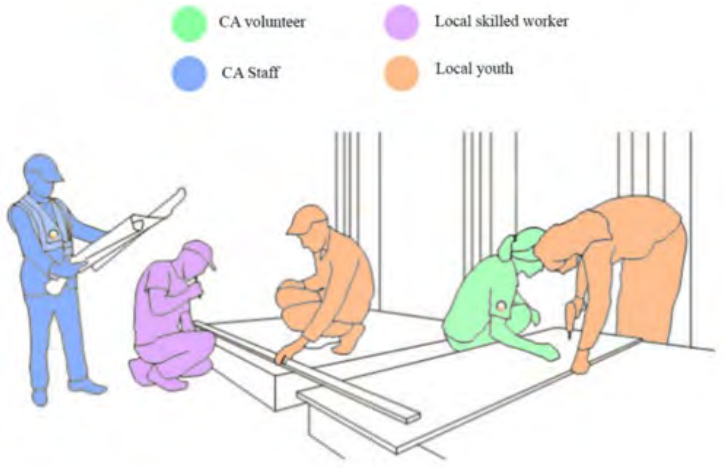


Figure 16. *This graphic depicts, by way of illustration, the various people (CatalyticAction staff and volunteers, local youth, and local skilled and semi-skilled workers) involved in the construction of the Jarahieh School (source: modified by author based on an illustration provided in "Pavilion Re-Claimed in Lebanon"). The construction was a collaborative effort that involved providing job opportunities for the locals and teaching locals especially local youth new skills in construction, which instilled a sense of empowerment within the community.*

9. Assessing the Adaptive Reuse Potential of Temporary Structures in Humanitarian Crises

9.1 Dearth of Adaptive Reuse Literature and Research in the Humanitarian Field

Because adaptive reuse literature and research studies have mainly concerned urban revitalization and commercial development in metropolitan areas, there

[49] Ariana Zilliacus, "With the Jarahieh Refugee School, CatalyticAction Demonstrates the True Potential of Temporary Structures," *ArchDaily*, March 2, 2017.
[50] LafargeHolcim"Pavilion Re-Claimed in Lebanon," Holcim Foundation.

seems to be a very little scholarship or investigations that have been carried out in regard to adaptive reuse potential for humanitarian purposes or in the context of addressing a humanitarian crisis. Exceptionally, several studies have examined the adaptive reuse potential of old, vacant buildings as temporary shelters for refugees.[51] However, these studies appear to focus mainly on assessing whether the built environment can be converted to meet the immediate housing needs of refugees and do not weigh in factors that may be most relevant from the beneficiaries' perspective, such as their social-physical well-being.

To help deal with the ongoing global refugee crisis, a project called Better Shelter was showcased in 2016 as a collaboration between IKEA and the United Nations Human Rights Council.[52] The project involved the manufacture and provision of modular shelters made from an adjustable frame for refugee camps across various parts of the world.[53] But while such a project is honorable in helping to address the immediate health and safety of refugees and forcibly displaced persons, it is not a sustainable mid- or long-term solution to meet the needs of such individuals in terms of overall quality of life, such as their comfort, well-being, and other social dimensions like education of young children who most need attention and care.

Hence innovative, inventive solutions and ideas beyond merely providing temporary, emergency shelters will need to be continuously studied and researched in the humanitarian field.[54] While temporary, emergency shelters may be rapidly deployable and cost-effective in addressing minimum survival conditions, they are not a mid or long-term viable solution in terms of addressing the basic social-physical needs of people suffering from a humanitarian crisis.

9.2 Adaptive Reuse Potential in Dealing with Humanitarian Crises

The Jarahieh School project is an innovative example of adaptive reuse as it entailed (i) changing the function of an existing building structure (from a temporary exhibition structure to a multi-purpose school) and (ii) recovering and reusing substantially all the materials and components of the original structure in the construction of the new structure. In the field of humanitarianism, there is great potential for developing and implementing projects based on applying these two adaptive use principles, beyond merely providing temporary, emergency shelters as the primary solution, to meet the social-physical as well as other basic human needs of refugees and displaced persons.

More importantly, the Jarahieh school project is exemplary in that, in

---

[51] Haniyeh Razavivand Fard and Asma Mehan, "Adaptive Reuse of Abandoned Buildings for Refugees: Lessons from European Context," *Suspended Living in Temporary Space: Emergencies in the Mediterranean Region: International Conference Proceeding*, 189–97, LetteraVentidue, 2018; Carla Bruni, "Vacant Buildings for Refugees: A Case Study in the Power of Adaptive Reuse of Older and Historic Buildings for Resilience," World Heritage USA, United States Committee of the International Council on Monuments and Sites.

[52] Dima Stouhi, "Architectural Responses to Humanitarian Crises Beyond Designing Buildings," *ArchDaily*, February 9, 2022.

[53] Ibid.

[54] Christele Harrouk, "Refugee Camps: From Temporary Settlements to Permanent Dwellings," *ArchDaily*, July 26, 2021.

responding to a humanitarian crisis, it utilized three approaches: participatory design, sustainable design, and community-led construction. These approaches, of many forms in different contexts, have been practiced and studied in various fields and areas of architecture and urban planning, but they have not been given serious attention by academia and practitioners.

One of the reasons may be due to the generalized understanding that refugee crises are temporary or impermanent phenomena and, therefore, temporary, emergency shelters may be adequate as a solution. However, in reality, ongoing refugee crises in various parts of the world in recent years show that such understanding may not be true as there are millions of refugees and displaced persons who have lived persistently in humanitarian crisis conditions on a long-term basis.

To help address the ongoing refugee crises, the potential solution cannot be limited to discussions at the level of how best to provide and meet the immediate housing needs of such persons. Rather, refugees' particular needs must be carefully taken into account, their voices must be heard, and they should be invited to partake and be allowed to contribute to the very process of ideating, designing, and implementing the potential solution(s) to address their social-physical as well as other basic human needs.

The three approaches utilized in the Jarahieh School project serve as an ideal framework for how sustainable adaptive reuse can potentially be applied effectively in projects involving the use of temporary structures in humanitarian crises.

10. Conclusion

By exploring the conception of ideas, design and implementation of Jarahieh School project, this paper highlights how adaptive reuse can be inventively applied to contexts that are not considered by many as traditional or typical. One such context is the ongoing humanitarian crisis.

Refugee settlements and other locations where displaced persons have been living have been constructed using cost-efficient, immediate tents aimed to address their short-term housing needs. However, the reliance on these structures alone may have shortcomings. They may be unfair and not humane to the people living at these sites and for their future generations.

In terms of significance, the Jarahieh School project should not be viewed as just a creative instance of adaptive reuse, but rather it represents a potential for providing a more permanent and site-specific solution to help deal with the ongoing humanitarian crisis.

Through a case study of the Jarahieh School project, this paper puts forward an ideal framework for assessing the adaptive reuse potential of temporary structures to respond to humanitarian crises. Participatory design, sustainable design, and community-led construction are the three approaches utilized in designing and implementing the Jarahieh School project. The participatory design approach is based on the principle of social inclusion and collaboration with the beneficiaries and other stakeholders during the idea conception and design phrases of the project. The sustainable design approach rests on recycling and reusing temporary structures which would otherwise be

discarded or made obsolete and makes use of them as construction materials while adapting to the local context. The community-led construction approach seeks to achieve the goal of installing a sense of ownership and empowerment on the part of the beneficiaries as well as the local community at large by engaging them directly in the project's construction and implementation phases. Adaptive reuse of temporary structures based on applying the three approaches can potentially be a viable means of improving the welfare of the people suffering in humanitarian crises.

Potential contributions to the practice of architecture and more specifically in the field of sustainable adaptive reuse could result from a better understanding of how the three approaches can be used effectively in adaptive reuse projects in humanitarian crisis conditions.

Bibliography

Berlanda, Toma. "*2019 On Site Review Report: Jarahieh School.*" Accessed on July 20, 2022. https://s3.us-east-1.amazonaws.com/media.archnet.org/system/publications/contents/14047/original/DTP106431.pdf?1586950598.

Blander, Akiva. "Diamond Schmitt Architects Adapts a Historic Train Station for the Canadian Senate." *Metropolis*. March 29, 2019. https://metropolismag.com/projects/diamond-schmitt-architects-canadian-senate.

Bruni, Carla. "Vacant Buildings for Refugees: A Case Study in the Power of Adaptive Reuse of Older and Historic Buildings for Resilience." World Heritage USA. United States Committee of the International Council on Monuments and Sites. Accessed July 23, 2022. https://worldheritageusa. org/vacant-buildings-for-refugees-a-case-study-in-the-power-of-adaptive-reuse-of-older-and-historic-buildings-for-resilience.

Bullen, Peter A., and Peter E.D. Love. "Adaptive Reuse of Heritage Buildings." *Structural Survey* 29, no. 5 (2011): 411–21. https://doi.org/10.1108/02630801111182439.

"CatalyticAction: Jarahieh School for Syrian Refugee Children in Lebanon." Floornature Architecture & Surfaces. June 11, 2019. https://www.floornature.com/catalyticaction-jarahieh-school-syrian-refugee-children-leba-15034.

Crook, Lizzie. "Atelier Liu Yuyang Reuses Old Farmhouses to Create Boutique Hotel in Rural China." dezeen. Dezeen Limited, October 15, 2019. https://www.dezeen.com/2019/10/14/xy-yunlu-hotel-atelier-liu-yuyang-architecture-china.

Debaj, Joana, Ricardo Conti, Matteo Zerbi, Elena Brunete, and Ronan Glynn. "Pavilion Re-claimed: Adaptive Reuse for Refugee Education, El Marj, Lebanaon (Bronze Award 2017)." Holcim Foundation. Accessed August 1, 2022. https://src.holcimfoundation.org/dnl/29bd45 09-c300-44e9-bd05-4b549f2561db/A17MEAbrLB.pdf.

De Carli, Beatrice, Celia Macedo, and Lucia Caistor-Arendar. "CatalyticAction: Interview of Joana Dabaj." *Pedagogies of Inclusion Vol.1: A review of Spatial Design Education in Europe (2019).* University of Sheffield (2019). https://issuu.com/asf-uk/docs/839713_6793cfb3c7b14a34bef8f7fda1b96294.

Dimitrova, Daniela V., Emel Ozdora-Aksak, and Colleen Connolly-Ahern. "On the Border of the Syrian Refugee Crisis: Views from Two Different Cultural Perspectives." *American Behavioral Scientist* 62, no. 4 (February 2018): 532–46. https://doi.org/10.1177/0002764218756920.

Forsey, Helen. "Excerpt: Envisaging a People's Senate." Canadian Centre for Policy Alternatives. April 1, 2015. https://policyalternatives.ca/publications/monitor/excerpt-envisaging-peoples-senate.

"Ghirardelli Square." The Landscape Architecture of Lawrence Halprin (The Cultural Landscape Foundation). Accessed July 26, 2022. https://www.tclf.org/sites/default/files/microsites/halprinlegacy/ghirardelli-square.html.

Goel, Eeti. "Adaptive Reuse Architecture: Breathing New Life in Structures." The Design Gesture. Accessed July 28, 2022. https://thedesigngesture.com/adaptive-reuse-architecture.

Harrouk, Christele. "Refugee Camps: From Temporary Settlements to Permanent Dwellings." ArchDaily. July 26, 2021. https://www.archdaily.com/940384/refugee-camps-from-temporary-settlements-to-permanent-dwellings.

Heilmeyer, Florian. "6 Projects That Made the Netherlands a World Capital of Adaptive Reuse." Metropolis, February 17, 2021. https://metropolismag.com/projects/netherlands-adaptive-reuse.

"Jarahieh School." CatalyticAction. Accessed August 23, 2022. https://www.catalyticaction.org/jarahieh-school.

"Jarahieh School, Al-Marj, Lebanon." Archnet. Accessed July 26, 2022. https://www.archnet.org/sites/19007.

Knaub, Steve. "The Potential of Adaptive Reuse." Bill Gladstone Group. NAI Commercial-Industrial Realty Co., August 14, 2020. https://www.billgladstone.com/publication-articles/the-potential-of-adaptive-reuse.

Langston, Craig, Francis K.W. Wong, Eddie C.M. Hui, and Li-Yin Shen. "Strategic Assessment of Building Adaptive Reuse Opportunities in Hong Kong." *Building and Environment* 43, no. 10 (2008): 1709–18. https://doi.org/10.1016/j.buildenv.2007.10.017.

Larkham, Peter J. "Rebuilding the Industrial Town: Wartime Wolverhampton." *Urban History* 29, no. 3 (2002): 388–409. https://doi.org/10.1017/s0963926802003048.

Owojori, Oluwatobi, Chioma Okoro, and Nicholas Chileshe. "Current Status and Emerging Trends on the Adaptive Reuse of Buildings: A Bibliometric Analysis." *Sustainability* 13, no. 21 (October 21, 2021): 11646. https://doi.org/10.3390/su132111646.

"Pavilion Re-Claimed in Lebanon." Holcim Foundation. Accessed August 1, 2022. https://src.lafargeholcim-foundation.org/flip/A18/Pavilion-Reclaimed.

Razavivand Fard, Haniyeh, and Asma Mehan. “Adaptive Reuse of Abandoned Buildings for Refugees: Lessons from European Context .” Essay. In *Suspended Living in Temporary Space: Emergencies in the Mediterranean Region: International Conference Proceedin*, 189–97. LetteraVentidue, 2018. https://philarchive.org/archive/FARARO-4.

“Repurposed Exposition Pavilion.” Piggy Backing Practices (a virtual symposium) of University of Arkansas, School of Architecture + Design. https://piggybackingpractices.com/expo-pavilion.

Shahi, Sheida, Mansour Esnaashary Esfahani, Chris Bachmann, and Carl Haas. “A Definition Framework for Building Adaptation Projects.” *Sustainable Cities and Society* 63 (December 2020): 102345. https://doi.org/10.1016/j.scs.2020.102345.

Stouhi, Dima. “Architectural Responses to Humanitarian Crises Beyond Designing Buildings.” *ArchDaily*. February 9, 2022. https://www.archdaily.com/976502/architecture-philanthropy-and-the-responses-to-humanitarian-crises-beyond-designing-buildings.

“Temporary Expo Structure Is Repurposed as School for Syrian Refugees in Lebanon.” *REVITALIZATION: The Journal of Urban, Rural & Environmental Resilience.* September 17, 2017. https://revitalization.org/article/temporary-expo-structure-repurposed-school-syrian-refugees-lebanon.

“XY Yunlu Hotel by Atelier Liu Yuyang: Dezeen Awards (Shortlist for Hospitality Building of the Year).” Dezeen. Dezeen Limited. Accessed July 28, 2022. https://www.dezeen.com/awards/2019/shortlists/xy-yunlu-hotel.

Zilliacus, Ariana. “With the Jarahieh Refugee School, CatalyticAction Demonstrates the True Potential of Temporary Structures.” *ArchDaily*. March 2, 2017. https://www.archdaily.com/806427/with-the-jarahieh-refugee-school-catalyticaction-demonstrates-the-true-potential-of-temporary-structures.

Animal Images in Christian Art of the Medieval Mediterranean: Perceptions of the Cultural "Other"

Fengyi Han

Author Background: *Fengyi Han grew up in China and currently attends Blair Academy in Blairstown, New Jersey, in the United States. Her Pioneer research concentration was in the field of art history and was titled "The Medieval Mediterranean: Confluence of Cultures."*

Abstract

This paper explores how specific animal images appearing in Christian artworks in the medieval Mediterranean have been isolated from the main artwork as the presence of a cultural "other," specifically referring to Islamic culture. The artworks in question originate from locations of intense interactions between Islam and Christianity such as Sicily and Spain, respectively. In the case of the Coronation Mantle of Roger II, scholars have interpreted the imagery of the defeated camel by the lion to represent the subjugation of Muslims under Norman Christians. The frescoes in San Baudelio de Berlanga have been divided by scholars' interpretations between the "Islamic" animal frescoes on the ground floor and the frescoes of Christian religious scenes on the second level. This paper argues that the significance of these animal images is more nuanced and historically specific than simply a statement of Muslim versus Christian power and proposes alternative interpretations.

Introduction

"[W]e must take into account the myriad ways in which animals, wild and domesticated, are entwined in human cultural history: animals, after all, are foes and friends, symbols and signs"

Thomas T. Allsen, *The Royal Hunt in Eurasian History*[1]

[1] Thomas T. Allsen, The Royal Hunt in Eurasian History, Encounters with Asia (Philadelphia: University of Pennsylvania Press, 2006), 10, quoted in Sharon Kinoshita, "Animals and the Medieval Culture of Empire," in Animal, Vegetable, Mineral, ed. Jeffrey Cohen (Washington, DC: Oliphaunt Books), 37, https://www.academia.edu/1874891/Animals_and_the_Medieval_Culture_of_Empire.

The origins of portable objects (works that can be moved or carried as personal objects) in the medieval Mediterranean are often hard to pinpoint due to the variety of cultures and the cultural fluidity during that time. Eva Hoffman termed this network of objects crossing physical and cultural borders in the medieval Mediterranean the "pathways of portability."[2] Along these pathways, animal imagery also played a part in the cross-cultural interactions between different cultures and religions in addition to physical objects. In order to more accurately understand what these images and objects reveal about the cross-cultural interchanges that had occurred and the people behind the exchanges, scholars are encouraged to reconceptualize their approaches. Multicultural works are traditionally discussed within geographical, cultural, or religious boundaries. However, such unidimensional labelings often do not comfortably fit objects from the medieval Mediterranean with dynamic origins and inconstant environments. Especially when analyzing works with seemingly both Christian and Islamic attributes, there is a tendency for scholars to interpret Islamic aspects separately from the Christian ones and to frame Islam as a cultural "other."[3] For example, animals such as camels were not native nor commonly used in some parts of the Mediterranean like Spain and Sicily. As a result, when renderings of such animals appear in Christian artworks originating from the aforementioned locations, they tend to stand out within the overall decorative scheme. Often, scholars tend to interpret these "non-Western" animals through a predominantly Western lens as markers of a foreign culture.

In the case of the Coronation Mantle of Roger II (1133), a king of Norman Sicily, its embroideries show camels being crushed under lions (fig. 1 Mantle). The camels have been interpreted as symbols of Islam by previous scholars and the scene as representing the conquest of Muslim Sicily and its Arab occupants by the Normans.[4] Comparably, the prominent animal frescoes in San Baudelio de Berlanga in Spain (early 12th century) have often been discussed separately from the church's other frescos that feature Christian religious scenes (fig. 2 camel) (fig. 5 animals); scholars have interpreted the animal images as an appropriation of Islamic art by reconquering Christians.[5] Both of these artworks come from areas of intense cultural interactions in the medieval Mediterranean. Sicily by nature is the crossroad of the Mediterranean Sea as it is connected to the people of North Africa, such as the Ifriqiya people and the Fatimid Caliphate in Egypt, and the Byzantine Empire in the East by water.[6] As a result, Sicily is a key location

---

[2] Eva R. Hoffman, "Pathways of Portability: Islamic and Christian Interchange from the Tenth to the Twelfth Century," *Art History* 24, no. 1 (2007): 320, https://doi.org/10.1111/1467-8365.00248.

[3] See the interpretation of the camel imagery on Roger II's mantle in Hoffman, "Pathways of Portability," 326-329, and the interpretation of frescoes in San Baudelio de Berlanga in Jerrilynn D. Dodds, "Hunting for Identity," in *Imágenes y promotores en el arte medieval: miscelánea en homenaje a Joaquín Yarza Luaces,* ed. Ma Luisa Melero Moneo (Bellaterra: Servei de Publicacions de la Universitat Autònoma de Barcelona, 2001), 97. Google Books.

[4] Hoffman, "Pathways of Portability," 328.

[5] Dodds, "Hunting for Identity," 97.

[6] Fein, Ariel, "The visual culture of Norman Sicily," Khan Academy, accessed August 9, 2022, https://www.khanacademy.org/humanities/medieval-world/byzantine1/x4b0eb531:middle-byzantine/a/the-visual-culture-of-norman-sicily

to control due to its connections to Europe, Africa, and Asia. In the eleventh century, Norman conquerors overtook southern Italy and Sicily from the previous Byzantine and Islamic populations during their conflict.[7] The Normans unified the entire region as the Kingdom of Norman Sicily, and under the rule of Norman kings Roger II (r. 1130-1154), William I (r. 1154-1166), and William II (r. 1166-1189), Sicily thrived as a multicultural and multi-ethnic place.[8] On the other hand, the medieval Iberian peninsula (modern Spain) was marked by continual conflict between Al-Andalus under Islamic control and the reconquering Christians in the north. The location of San Baudelio Berlanga in Soria is especially close to the frontier of conflict (fig. 3 map) at the time of its construction and fresco decoration.[9] If the specific historical contexts of Sicily during the Mantle's production and the context of other frescoes in San Baudelio de Berlanga are considered, alternative readings of the animal imagery may emerge. This paper argues that the inclusion of animals in art located in Islamic-Christian frontier zones, such as the camel and lion imagery on the Mantle of Roger II and the animal frescoes in San Baudelio de Berlanga, is more nuanced and historically specific than simply a statement of Muslim versus Christian power.

Coronation Mantle of Roger II

One object which provides further insight into the aforementioned problem is the Coronation Mantle attributed to Roger II, the Norman king of Sicily from 1130 to 1154 (fig. 1 Mantle).[10] Following the Normans' conquest of Sicily and southern Italy from Arab rulers in 1060, the Normans began to cultivate a culture in Sicily that included both Christian components from the Latin West and components from Byzantine and Islamic cultures.[11] The Mantle made 3 years after his coronation captures this multicultural aspiration of Roger II in its design. Measuring 143 cm tall and 345 cm wide, the Mantle and its red silk, gold thread embroideries, and luxurious embellishments made of pearls and enamels are surprisingly well preserved.[12] The Mantle has a semi-circular shape, which may be of Latin origin, as similar hemispherical mantles were used by monarchs such

[7] Fein, Khan Academy, "The visual culture of Norman Sicily."

[8] For dates of kingship, see John Julius Cooper, "Roger II," in *Encyclopedia Britannica*, February 22, 2022, https://www.britannica.com/biography/Roger-II;
The Editors of Encyclopaedia Britannica, "William I," in *Encyclopedia Britannica*, 3 May. 2022, https://www.britannica.com/biography/William-I-king-of-Sicily;
The Editors of Encyclopaedia Britannica. "William II". *Encyclopedia Britannica*, 1 Jan. 2022, https://www.britannica.com/biography/William-II-king-of-Sicily.
For thriving Sicily, see Khan Academy, "The visual culture of Norman Sicily."

[9] Lauren Kilroy-Ewbank and Steven Zucker, "Camel from San Baudelio de Berlanga," Smarthistory, May 5, 2020, https://smarthistory.org/camel-spain/.

[10] John Julius Cooper, "Roger II," in *Encyclopedia Britannica*, February 22, 2022, https://www.britannica.com/biography/Roger-II.

[11] Hoffman, "Pathways of Portability," 325–326.

[12] Steven Zucker and Beth Harris, Smarthistory, "Coronation Mantle."

as the Holy Roman Emperor Henry.[13] It also shows influences of Islamic-Byzantine art styles since gifts of semicircular silk cloaks with decorations of lions or other animals given by the Byzantine emperor to the Caliph have been recorded.[14] Kufic script is embroidered around the edge of the Mantle and states that:

> This is what was made in the royal treasury (khizanah). Full happiness, honour, good fortune, perfection, long life, profit, welcome, prosperity, generosity, splendour, glory, perfection, realization of aspirations and hopes, of delights of days and nights, without end or modification, with might, care, sponsorship, protection, happiness, well-being (success), triumph and sufficiency. In Palermo (Madinah Siquliyah) in the year 528 [1133–4].[15]

The Arabic inscription and the multiple possible sources for the shape of the mantle make this work a classic example of the multiculturalism present in the art of the medieval Mediterranean. The eye-catching animal imagery has been a particular point of focus for scholars attempting to analyze its meaning.

Animal Imageries on the Mantle

Sewn along the vertical axis of the Mantle is a palm tree, and reflected on either side is the imagery of a lion subduing a camel (fig. 4 Mantle detail). Lions are a common symbol of kingship, bravery, and strength in both Christian and Islamic cultures.[16] The choice to display the lion in combat underscores the notion of power shown by the iconography and emphasizes the theme of dominance and submission. The combat scene can be assumed to praise Normans' victories over adversaries in general.[17] Additionally, the lion could be a characterization of Roger II himself, as the lion was a symbol of royalty in Norman royal monuments, and he had previously used the image of a victorious lion to symbolize himself.[18]

Ambiguities begin to arise when determining the meaning of the camel under the claws of the lion on the Mantle. The animal combat scene, more commonly seen as a lion attacking a bull, has been a common motif of power

---

[13] Clare Vernon, "Dressing for Succession in Norman Italy: The Mantle of King Roger II," *Al-Masāq* 31, no. 1 (January 2, 2019): 3, https://doi.org/10.1080/09503110. 2018. 1551699.

[14] *Book of Gifts and Treasures: Selections Compiled in the Fifteenth Century from an Eleventh-Century Manuscript on Gifts and Treasures (kitaʾb Al- hadāyāʾ Wa Al-tuḥaf),* trans. Ghada Hijjawi Qaddumi (Cambridge, MA: Harvard University Press 1996), 262–263.

[15] Hoffman, "Pathways of Portability," 329.

[16] Mirjam Gelfer-Jørgensen, *Animal Motifs,* trans. Caroline C. Henriksen, (Leiden: Brill Academic Publishers, 1986), 111–39. Google Books.

[17] Clare Vernon, "Dressing for Succession in Norman Italy," 2.

[18] Hoffman, "Pathways of Portability," 327

circulating since the Achaemenid period in Persian Art.[19] The choice of replacing the bull with a camel on the Mantle of Roger II is a deliberate one and has warranted much speculation regarding its significance. A popular reading proposed by Eva Hoffman and other scholars sees the camel as a symbol of Muslim defeat by the Norman Christian.[20] Eva Hoffman in "Pathways of Portability" speculates that camels, a species exotic to Sicily, can only have been used to represent the Muslim population. She also cites Oleg Grabar's "Experiences of Islamic Art" as evidence of similar utilizations of the camel iconography in Norman and Crusader contexts to represent foreign cultures. Indeed, it seems logical that Roger II would express his triumph over the newly conquered land from the Muslim Sicilians with his Coronation Mantle. However, Grabar himself said in the same article that the association of camel imagery with Muslims is purely anachronistic as there are no medieval sources, Eastern or Western, that connect camels to Muslims or Arabs at the time of the Mantle's creation.[21] Thus, speculation of the camel representing Islam is unsubstantiated, and this interpretation, if it is not influenced by this notion itself, propels the inaccurate view of Islam being viewed as the inferior 'other' in places considered Western.

Additionally, a mantle celebrating vanquished Muslims would have been inconsistent with the reality of Norman Sicilian culture and Roger II's political mentality. Even before conquering Sicily entirely, Roger I, the father of Roger II, had incorporated a considerable number of Muslim soldiers into his forces. After completing his conquest, Roger I never forced the people of Islamic faith to convert to Christianity.[22] Political capitals such as Messina and Palermo also had a great deal of cultural assimilation between Christians and Muslims. The account of a pilgrim to Mecca, Ibn Jubayr, who stayed in Palermo in 1184, described how Christian women were dressed as they headed towards Christmas Mass in the Church of la Martorana:

> The Christian women's dress in this city is the dress of Muslims; they are eloquent speakers of Arabic and cover themselves with veils. They go out at this aforementioned festival clothed in golden silk, covered in shining wraps, colorful veils and with light gilded sandals. They appear at their churches bearing all the finery of Muslim women in their attire, henna and perfume. [23]

---

[19] Vijay Sathe, "The Lion-Bull Motifs of Persepolis: The Zoogeographic Context." *Iranian Journal of Archaeological Studies* 2, no. 1 (January 1, 2012): 75. https://ijas.usb.ac.ir/article_1059_c652b8753b1dba4106e5d75c5d106b6b.pdf

[20] Hoffman, "Pathways of Portability," 328.

[21] Oleg Grabar, "The Experience of Islamic Art," in *The Experiences of Islamic Art on the Margins of Islam*, ed. Irene A. Bierman, (Berkshire: Garnet Publishing Limited, 2005), 37. Google Books.

[22] Karen C. Britt, "Roger II of Sicily: Rex, Basileus, and Khalif? Identity, Politics, and Propaganda in the Cappella Palatina," *Mediterranean Studies* 16 (2007): 23. https://www.jstor.org/stable/41167003.

[23] Ibn Jubayr, *Rihlat Ibn Jubayr*, (Beirut: Dar al-Sadir 1962), 307, trans. A. Metcalfe, *Muslims and Christians in Norman Sicily: Arabic-Speakers and the End of Islam*, 97, quoted in Britt, "Roger II of Sicily," 25.

Evidently, the attitude towards the different religions and cultures in the cosmopolitan Kingdom of Norman Sicily is rather accepting. Likewise, Roger II was educated by Greek and Muslim teachers, fluent in Arabic, greatly invested in upholding a multicultural court, and had a generally tolerant attitude toward his Muslim subjects.[24] As a result, it is evident that Muslims were not viewed exclusively as a religious adversary by the Normans as Eva Hoffman and other scholars previously suggested in their interpretations of the animal imagery on the Mantle. Such an accommodating attitude makes a mantle celebrating Muslim defeat unnecessary and even counterintuitive, as it could spark resentment and revolts in the predominantly Muslim population of Sicily, which is a reaction Roger I feared during his reign.[25]

But if the animal imagery does not represent Normans' conquest of Muslims, what could it mean? A possible connection may be with Africa, since the late Roman Empire did identify camels with desert nomads.[26] This fact brings to light another possible analysis of the animal imagery on the Mantle. During Roger II's reign, Normans had wished to conquer Ifriqiya in hopes of expanding their power and wealth. The Normans eventually achieved this goal in the 1140s, when droughts devastated Ifriqiya.[27] Thus, the camels on the Mantle may be a reference to the native nomads of Ifriqiya (modern-day Tunisia) in Northern Africa, as they used these desert beasts of burden.[28] Considering the historical context of Norman Sicily during the reign of Roger II, the lion clutching the throat of the camel proclaims the message that Roger II's power will lead him to successful conquests against anyone that stands in the way of the Normans.

Although this interpretation challenges the simple viewpoint of Christian powers versus the "foreign" Muslim forces, the camel imagery is still being thought of as a representation of a cultural "other." Thus, it is necessary to consider if the animal iconography on Roger II's Mantle also had a place in the Norman Christian context. William Tronzo, in "The Mantle of Roger II of Sicily," notes that camels were often used to carry corrupt or usurping leaders in public as a form of humiliation. For example, when the Byzantine emperor Andronikos I Komnenos was captured for crimes such as conspiring against Emperor Manuel I Komnenos in the twelfth century, he was paraded through Constantinople on the back of a camel while being beaten by the public; when anti-pope Gregory VIII was turned over to the legitimate pope Calixtus II in 1121, he was similarly

---

[24] Timothy Smit, "Pagans and Infidels, Saracens and Sicilians: Identifying Muslims in the Eleventh-Century Chronicles of Norman Italy," in *The Haskins Society Journal 21: Studies in Medieval History*, ed. William L. North (Boydell & Brewer, 2010), 83-84, https://www.academia.edu/474443/Pagans_and_Infidels_Saracens_and_Sicilians_Identifying_Muslims_in_the_Eleventh_Century_Chronicles_of_Norman_Italy.

[25] Britt, "Roger II of Sicily," 23.

[26] Grabar, "The Experience of Islamic Art," 37.

[27] Matt King, *Dynasties Intertwined: The Zirids of Ifriqiya and the Normans of Sicily* (New York: Cornell University Press, 2022), 138. Google Books.

[28] Robin S. Reich, "The Lion and the Camel: The Mantle of Roger II and Siculo-Norman Relations with the Islamicate Mediterranean," 15, https://www.academia.edu/37614022/The_Lion_and_the_Camel_The_Mantle_of_Roger_II_and_Siculo_Norman_Relations_witht_the_Islamicate_Mediterranean_docx.

paraded through Rome seated backward on a camel.[29] As a result, the camel could have been viewed as a symbol of incompetent and illegitimate rulers in the eyes of Normans. The imagery of the lion overcoming the camel on the Mantle acts as a reminder of the consequences of misusing authority. Thus, the animal imagery reminds Roger II and his subject to stand above abusing royal power and defeating those who do.

After consideration of the larger historical context of Norman Sicily and the medieval Mediterranean, the interpretation of the camel being crushed by the lion as representing Muslim subjugation under Norman rule is less substantiated, and alternative interpretations surface. It is important to note the phenomenon of scholars separating the image of the camel in Christian art from other imagery and interpreting it as a symbol of the "foreign" Islamic culture in locations where Christian-Islamic cultural interchange is prevalent. This method reflects the age-old tendency of Latin-Christian writers to construct a narrative of bipolar disparity with Muslims whilst both groups populated the Mediterranean.[30] The Mantle of Roger II demonstrates vividly how attempting to understand artworks in this period by dissecting them according to religious ideologies may cause scholars to overlook precisely the cultural complexities, interactions, and confluence that made up the medieval Mediterranean. The Mantle of Roger II has clear influences from both Islamic and Christian art styles, but that does not mean the coexistence of the two styles must exist in a hegemonic relationship. This mindset of the "foreign" and "inferior" Muslim is reflective of the modern negative perceptions of Islam in Western countries and the idea of northern Europe being the cultural "core" of the Mediterranean.[31] This misconception is a critical issue for scholars, as it can cloud the historical reality in the medieval Mediterranean when projected upon studies on such areas. Roger II's Mantle, no matter what the animal imagery actually represented, provides scholars with undeniable evidence that the Norman Sicily culture connected both the Latin-Christian culture and the Islamic culture into one. As a result, Norman Sicily cannot be fully understood if the Christian and Islamic aspects are alienated from each other, or restrictively labeled Islamicate, Byzantine, or Latin.

San Baudelio de Berlanga

Another set of artworks also sheds light on the problem of perceiving Muslim culture as a foreign presence and one that is incompatible with Christianity in the medieval Mediterranean: the frescoes in San Baudelio de Berlanga. The building

---

[29] William Tronzo, "The Mantle of Roger II of Sicily," in *Robes and Honor: The Medieval World of Investiture*, ed. Stewart Gordon (Palgrave, 2001), 249, https://www.academia.edu/9715402/_The_Mantle_of_Roger_II_of_Sicily_in_Robes_and_Honor_The_Medieval_World_of_Investiture_The_New_Middle_Ages_Series_ed_S_Gordon_New_York_and_London_Palgrave_2001_241_253.

[30] Jerrilynn D. Dodds, "Islam, Christianity, and the Problem of Religious Art," in *Late Antique and Medieval Art of the Mediterranean World* (Carlton, Australia: Blackwell Publishing, 2007), 350.

[31] Sarah Davis-Secord, Belen Vicens, and Robin Vose, *Interfaith Relationships and Perceptions of the Other in the Medieval Mediterranean: Essays in Memory of Olivia Remie Constable* (Cham: Springer Nature, 2021), 8.

is a small hermitage church constructed in Soria at the beginning of the 11th century [32], dedicated to St. Baudelius, a fourth-century martyr-missionary.[33] The area near the church was a frontier zone of the Christian reconquest of Muslim-controlled land (fig 3 map). The church is located only three miles away from Berlanga Del Duero, a major town that was under the rule of the Emirate of Zaragoza until the Christians successfully reconquered the land in 1037.[34] The church itself has a plain outer appearance, but the inside is richly decorated with vibrant frescoes, which were a later addition to the church, dating to the 12th century.[35] On the southern wall, there is a small entrance leading to a cavern, which is why scholars have ascribed the function of a hermitage to San Baudelio.[36] Most scholars agree that the church was constructed by Mozarab artists[37]; Mozarab refers to Spanish Christian people who lived under Muslim rule (8th-11th century) and adopted Muslim culture while not converting to Islam.[38] Upon entering the church through a horseshoe arched doorway located on the northern wall, the viewer can see remnants of frescos such as *The Hunter Pursuing Stag, Horseman with Hounds Chasing Hares*, and frescoes of animals such as a bear, elephant, etc. (fig. 5 animals), but the most striking work is the camel fresco facing the stairwell leading to the second floor (fig. 2 camel). The size of the fresco, 246 cm by 136 cm, is quite large, rendering the camel imagery physically larger and more eye-catching than the frescoes of Christian religious scenes on the higher level. The frescoes upstairs form a Christological cycle depicting events from the life of Christ such as *The Temptation of Christ* and the *Last Supper* as they stretch around all four walls (fig. 6 layout).

Interpretations of the frescoes in San Baudelio de Berlanga

Due to the disparate subject matters between the frescoes on the lower and upper floors, scholars have presented varying interpretations of this artistic phenomenon.[39] As a result of the prevalent conflicts between Christians and Muslims during this time period in Spain, interpretations from scholars often

---

[32] Annie Labatt, "Through the Eye of a Needle: Deciphering the Dromedary from San Baudelio de Berlanga," *Peregrinations: Journal of Medieval Art & Architecture* 5, no. 4 (2016): 42, https://digital.kenyon.edu/perejournal/vol5/iss4/2.

[33] Walter W. S. Cook, "Romanesque Spanish Mural Painting (II) San Baudelio De Berlanga," *The Art Bulletin* 12, no. 1 (1930): 22, https://doi.org/10.2307/3050760.

[34] Labatt, "Dromedary from San Baudelio," 46.

[35] Jerrilynn D. Dodds, "Hunting for Identity," in *Imágenes y promotores en el arte medieval: miscelánea en homenaje a Joaquín Yarza Luaces*, ed. Ma Luisa Melero Moneo (Bellaterra: Servei de Publicacions de la Universitat Autònoma de Barcelona, 2001), 97.

[36] Jerrilynn Denise Dodds, *Architecture and Ideology in Early Medieval Spain* (Penn State Press, 1990): 93.

[37] Cook, "Romanesque Spanish Mural Painting," 22.

[38] The Editors of Encyclopaedia Britannica, "Mozarab," in *Encyclopedia Britannica*, 2007, https://www.britannica.com/topic/Mozarab.

[39] See Dodds, "Hunting for Identity," 97-100, and Labatt, "Dromedary from San Baudelio," 45-46.

attribute the more secular animal frescoes on the first floor to a non-Latin Christian culture and discuss them separately from the frescoes of Christian scenes. For example, José Camón Azna wrote that the upper Christological cycle had a Romanesque style and the lower frescoes showed a Mozarabic style.[40] Because of this artistic difference, he suggests that a Muslim painter was responsible for the lower painting.[41] Jerrilynn Dodds in her "Hunting for Identity" points out that Juan Zozaya has linked every scene in the lower frescos with images that appeared on portable ivory objects from Al-Andalus.[42] As a result, Dodds argues that the inclusion of animals and hunting scene frescos in addition to Christian scenes appropriates these Islamic images and shows off Christian authority over the newly reconquered land around the church;

> If the animals can, on one level, allude to Islamic material culture taken on like booty, then the hunters provide the link to the newly conquered land by celebrating ownership in triumphant, monumental images. [...] [San Baudelio's design] converts ambivalence and artistic values shared with Islam into a bipolar opposition in which Islam is configured as other; [...]its history refocused on its apostolic past. [43]

Although Dodds' interpretation seems plausible, there are some factors to consider. Dodds herself dates the paintings in San Baudelio to 1143, which is more than a century after the Christian reconquest of Al-Andalus.[44] Would it make sense for an isolated hermitage to paint frescoes celebrating Christians' victory over Al-Andalus and Muslims after such a long time? Although Dodds' interpretation cannot be ignored, alternative interpretations can surface if we consider the animal frescoes with the Christian frescoes holistically instead of viewing the artworks through the paradigm of Muslim and Christian conflict.

One possible interpretation is that the animal frescoes were not used to appropriate Islamic culture. Instead, they actually work with the frescoes of Christian scenes above to form a distinct narrative. This may seem implausible at first; how can the fresco of a camel, an animal that does not physically exist in the Iberian Peninsula (not including the few imported ones), tie into the Christian frescoes in San Berlanga?[45] However, it is important to note that camels are also significant creatures in the Bible, especially when talking about the three magi.

> A multitude of camels shall cover you, the young camels of Mid'ian and Ephah; all those from Sheba shall come. They shall bring gold and frankincense, and shall proclaim the praise of the Lord.[46]

---

[40] Labatt, "Dromedary from San Baudelio," 45.

[41] Ibid.

[42] Dodds, "Hunting for identity," 92.

[43] Ibid., 98.

[44] Labatt, "Dromedary from San Baudelio," 47.

[45] Kilroy-Ewbank *et al*, Smarthistory, "Camel from San Baudelio de Berlanga."

[46] Isaiah 60:6.

According to the Bible, when the magi followed a star to Bethlehem in order to visit baby Christ and bring him gifts, they rode on camels as they journeyed from the East.[47] In San Baudelio, there is in fact a fresco of *The Adoration of the Magi* (fig. 7 magi) included in the frescoes of the Christological cycle on the second floor.[48] Although the fresco is missing pieces in its upper half due to damage by natural weathering and neglect, it is still possible to make out that the magi are riding three animals. The grey animal in the forefront and the black animal furthest back are horses, since the head of the black horse and the distinctive bushy tail of the grey horse are fully visible. The animal in the middle of the fresco appears to be a camel with its signature bright yellow color and long thin tail. Thus, this camel of the magus draws a comparison with the large camel fresco on the first floor.[49] To the viewer, it is almost as if the camel is leading visitors through a spiritual journey physically, as the view climbs the stairs of the church, and mentally, as the viewer follows the frescoes depicting the Christological cycle.[50] These frescoes, although seemingly indicative of different cultures, actually speak to each other if we deconstruct the artificially designated zones of "Christian" and "Muslim," the "other" and the familiar. As Labatt says in her article, "Dividing the frescoes stylistically is of questionable accuracy and ultimately fails to explain what is actually on the walls." Thinking of the upper and lower frescoes as a whole, one can connect the lower camel to the right of the stairs in the hermitage to the upper camels which appear in a Christian scene in a compelling narrative.[51] The camel guides the viewer through the church and strengthens the connection between the viewer and the holy scenes. Thus, the frescoes of San Baudelio call to attention that although these animal images show influences of Islamic art, that does not necessarily mean they represent Islamic culture.

Just as in the case of the Mantle of Roger II, the frescoes of San Baudelio cannot be fully understood within their cultural context if they are divided between the Christian aspects and the "Islamic." Interpretations using this methodology restrict scholars to view the frescoes at the hermitage merely as a product of Islamic and Christian conflict and perpetuate the binary construction that alienates Islamic history from the medieval Mediterranean. When we consider the historical background of the church and both levels of frescoes holistically, we can see that the story told by the artwork might not be one of Islamic subjugation under Christians after all.

---

[47] The Editors of Encyclopedia Britannica, "Magi," in *Encyclopedia Britannica*, 23 Nov. 2021, https://www.britannica.com/topic/Magi.

[48] Labatt, "Dromedary from San Baudelio," 52-53.

[49] Ibid., 55.

[50] Ibid.

[51] Ibid., 52.

Conclusion

"The past is a foreign country," says L. P. Hartley, "they do things differently there." [52] The medieval Mediterranean is a world characterized by fluidity, culturally and artistically. Scholars must be careful about the influence of modern perceptions of nationalism, Eurocentrism, or even Islamophobia on studies done on such a culturally interwoven place in time. We can better understand the reality of the medieval Mediterranean when we shed the misperception of Islam as a culture foreign to the Mediterranean and conduct contextually specific analysis. The Mantle of Roger II and the frescoes in San Baudelio de Berlanga are the perfect evidence of this fact. Examining past interpretations of the animal imagery on the Mantle of Roger II and the frescoes in San Baudelio Berlanga, while providing alternative ones based on the complex historical context of the art works, helps us understand the interchange between cultures in the medieval Mediterranean world that may otherwise be misunderstood as conflicting. This paper is not saying that medieval artworks from the Mediterranean never portrayed religious conflict or cultural appropriation. However, the idea of opposition, especially between Christianity and Islam, becomes problematic when applied without truly considering alternative interpretations or the historical background of the work. As new interpretations from additional historical perspectives shed light on the identity of cross-cultural artworks, they can offer a perspective on the complexity of identities in the medieval Mediterranean world beyond the simple paradigm of conflict between Christians and Muslims.

Illustrations

Figure 1. *Coronation Mantle of Roger II, Imperial Treasury, 1133. (Kunsthistorisches Museum, Vienna). www.khm.at/en/object/100435/*

[52] Leslie Poles Hartley, *The Go-Between* (London: Penguin Books, 1997), 5, quoted in David Blanks, "The Sense of Distance and the Perception of the Other," *Journal of Medieval Worlds* 1, no. 3 (September 3, 2019): 26, https://doi.org/10.1525/jmw.2019.130003.

Figure 2. *San Baudelio de Berlanga, Carmel Resco, Spain, Early 12th century. (Met Cloisters, New York). https://www.metmuseum.org/art/collection/search/471906*

Figure 3. *Map of Muslim-Christian frontier zone in the Iberian Peninsula around San Baudelio de Berlanga, 11th century. https://smarthistory.org/camel-spain/*

Figure 4. *Coronation Mantle of Roger II, detail of lion and camel image, Imperial Treasury, 1133. (Kunsthistorisches Museum, Vienna). www.khm.at/en/object/100435/*

Figure 5. *San Baudelio de Berlanga, reconstruction of San Baudelio de Berlanga showing the frescoes of the Elephant, the Bear, The Hunter Pursuing Stag, and Horseman with Hounds Chasing Hares from left to right, Spain, early 12th century. Photography by Félix González, 2016. https://flic.kr/p/Gng26N*

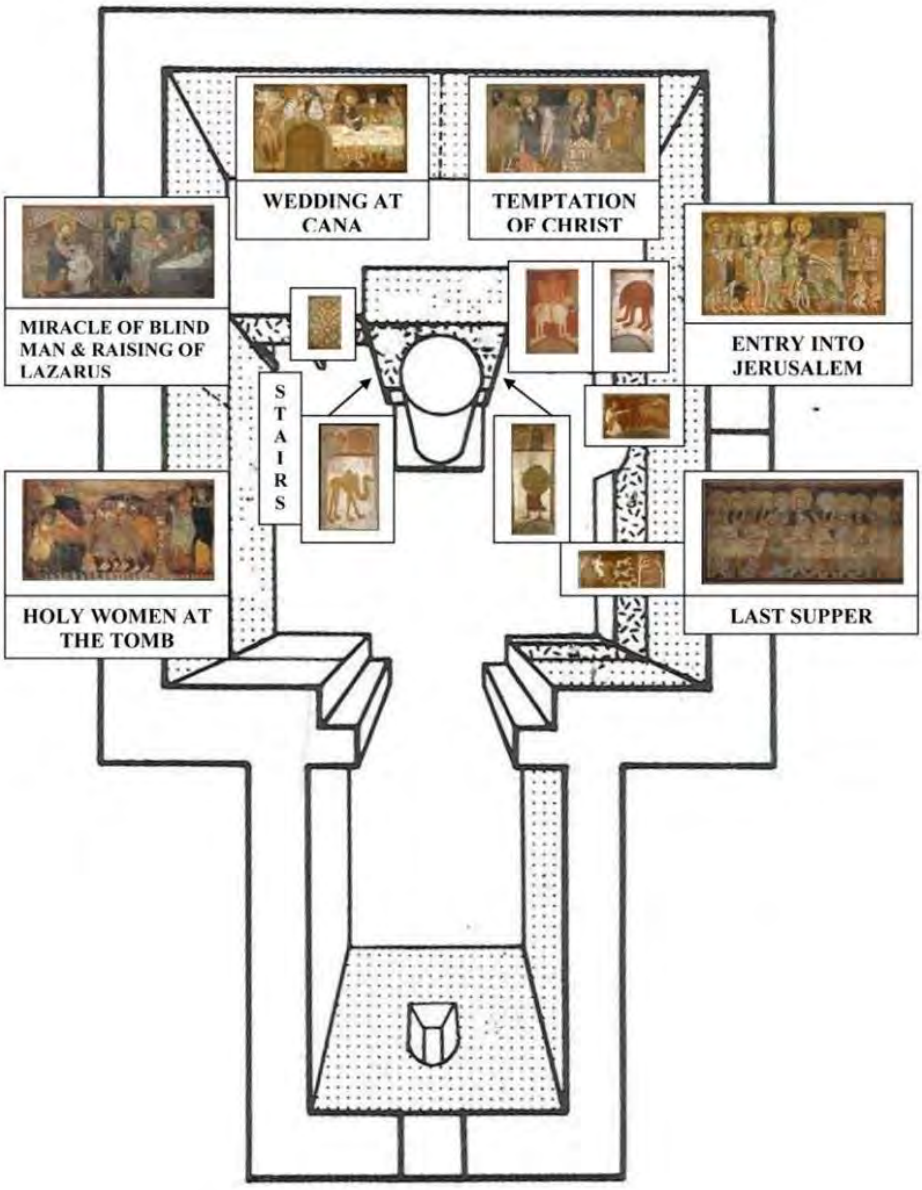


Figure 6. *San Baudelio de Berlanga, layout of the frescoes at San Baudelio de Berlanga, Spain, early 12th century. Diagram by Annie Labatt, 2016. https://digital.kenyon.edu/cgi/viewcontent.cgi?article=1040&context=perejournal*

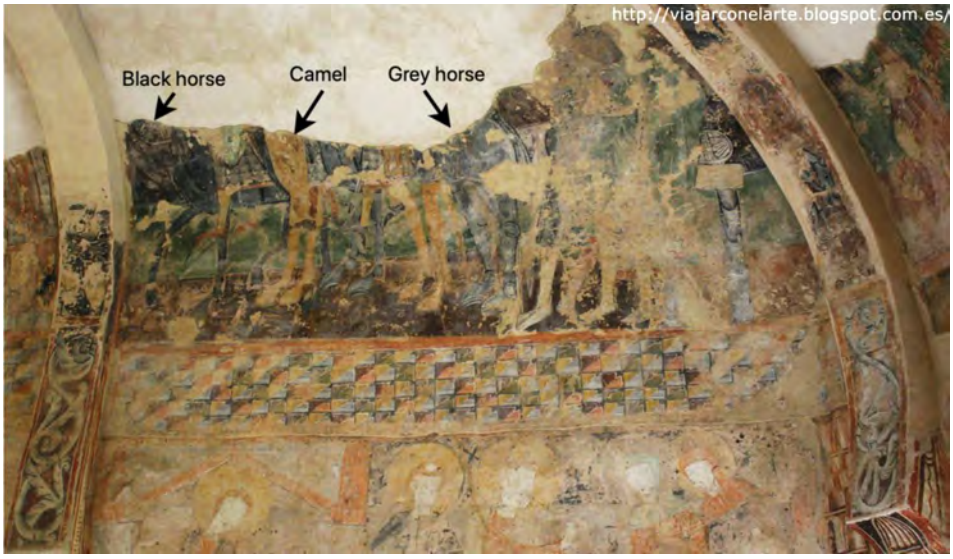


Figure 7. *San Baudelio de Berlanga, The Adoration of the Magi, Spain, early 12th century. Photograph by Sira Gadea, 2013. https://flic.kr/p/r2g3Z9*

References

Blanks, David. "The Sense of Distance and the Perception of the Other." *Journal of Medieval Worlds* 1, no. 3 (September 3, 2019): 21–44. https://doi.org/10.1525/jmw.2019.130003.

Britt, Karen C. "Roger II of Sicily: Rex, Basileus, and Khalif? Identity, Politics, and Propaganda in the Cappella Palatina." *Mediterranean Studies* 16 (2007): 21–45. https://www.jstor.org/stable/41167003

Cook, Walter W. S. "Romanesque Spanish Mural Painting (II) San Baudelio De Berlanga." *The Art Bulletin* 12, no. 1 (1930): 21–42. https://doi.org/10.2307/3050760.

Davis-Secord, Sarah, Belen Vicens, and Robin Vose. Interfaith Relationships and Perceptions of the Other in the *Medieval Mediterranean: Essays in Memory of Olivia Remie Constable*. Cham: Springer Nature, 2021. Google Books.

Dodds, Jerrilynn D. "Hunting for Identity." In *Imágenes y promotores en el arte medieval: miscelánea en homenaje a Joaquín Yarza Luaces*, edited by Ma Luisa Melero Moneo, 88-100. Bellaterra: Servei de Publicacions de la Universitat Autònoma de Barcelona, 2001. Google Books.

Dodds, Jerrilynn D. "Islam, Christianity, and the Problem of Religious Art," in *Late Antique and Medieval Art of the Mediterranean World*, edited by Eva R. Hoffman, 350-366. Carlton, Australia: Blackwell Publishing, 2007.

Grabar, Oleg. "The Experience of Islamic Art." *In The Experiences of Islamic Art on the Margins of Islam*, edited by Irene A. Bierman, 11–61. Berkshire: Garnet Publishing Limited, 2005. Google Books.

Gelfer-Jørgensen, Mirjam. *Medieval Islamic Symbolism and the Paintings in the Cefalù Cathedral*. Translated by Caroline C. Henriksen. Leiden: Brill Academic Publishers, 1986. Google Books

Hoffman, Eva R. "Pathways of Portability: Islamic and Christian Interchange from the Tenth to the Twelfth Century." *Art History* 24, no. 1 (2001): 17–50. https://doi.org/10.1111/1467-8365.00248.

King, Matt. *Dynasties Intertwined: The Zirids of Ifriqiya and the Normans of Sicily*. New York: Cornell University Press, 2022. Google Books.

Kinoshita, Sharon. "Animals and the Medieval Culture of Empire." In Animal, Vegetable, Mineral, edited by Jeffrey Cohen, 37–67. Washington, DC: Oliphaunt Books, 2012. https://www.academia.edu/1874891/Animals_and_the_Medieval_Culture_of_Empire.

Labatt, Annie. "Through the Eye of a Needle: Deciphering the Dromedary from San Baudelio de Berlanga." *Peregrinations: Journal of Medieval Art & Architecture* 5, no. 4 (2016): 41-63. https://digital.kenyon.edu/perejournal/vol5/iss4/2.

Reich, Robin S. "The Lion and the Camel: The Mantle of Roger II and Siculo-Norman Relations with the Islamicate Mediterranean." https://www.academia.edu/37614022/The_Lion_and_the_Camel_The_Mantle_of_Roger_II_and_Siculo_Norman_Relations_witht_the_Islamicate_Mediterranean_docx.

Sathe, Vijay. "The Lion-Bull Motifs of Persepolis: The Zoogeographic Context." *Iranian Journal of Archaeological Studies* 2, no. 1 (January 1, 2012): 75–85. https://ijas.usb.ac.ir/article_1059_c652b8753b1dba4106e5d75c5d106b6b.pdf

Smit, Timothy. "Pagans and Infidels, Saracens and Sicilians: Identifying Muslims in the Eleventh-Century Chronicles of Norman Italy." In *The Haskins Society Journal* 21: 2009 Studies in Medieval History, edited by William L. North, 67–86. Boydell & Brewer, 2010. https://www.academia.edu/474443/Pagans_and_Infidels_Saracens_and_Sicilians_Identifying_Muslims_in_the_Eleventh_Century_Chronicles_of_Norman_Italy.

Tronzo, William. "The Mantle of Roger II of Sicily." In *Robes and Honor: The Medieval World of Investiture*, edited by Stewart Gordon, 241-253. Palgrave, 2001. https://www.academia.edu/9715402/_The_Mantle_of_Roger_II_of_Sicily_in_Robes_and_Honor_The_Medieval_World_of_Investiture_The_New_Middle_Ages_Series_ed_S_Gordon_New_York_and_London_Palgrave_2001_241_253.

Vernon, Clare. "Dressing for Succession in Norman Italy: The Mantle of King Roger II." *Al-Masāq* 31, no. 1 (January 2, 2019): 95–110. https://doi.org/10.1080/09503110.2018.1551699.

Michelangelo's Aesthetic Philosophy: Discovering Divine Beauty

Khanh Nguyen

Author Background: *Khanh Nguyen grew up in Vietnam and currently attends Saigon South International School in Ho Chi Minh City, Vietnam. Her Pioneer research concentration was in the field of art history and titled "The Italian High Renaissance: Leonardo, Michelangelo, and Raphael."*

1. Introduction

Michelangelo Buonarroti (1475-1564) stood as an anomaly alongside other giants of the High Renaissance, for he, unlike many of his contemporaries and predecessors, mistrusted the application of mathematical models to achieve aesthetic perfection. While artists like Leonardo da Vinci (1452-1519) toiled and experimented with mathematical measures to find ideal proportions in art, Michelangelo insisted that the determinants of beauty are kept in the eyes, and that an artist's judgment alone can identify pleasing measures and proportions. [1] Michelangelo's aesthetic philosophy, then, centralizes around the idea that beauty should not be bound by set traditions and rules, and that aesthetic perfection can only be determined by the artist's abstract notions of beauty.

Michelangelo's aesthetic philosophy had driven him to create works in "sculpture, painting, and architecture that departed from High Renaissance regularity."[2] He claimed he did not owe his artistic virtuosity to anyone except God and his own genius, one cannot dismiss the influence of other artists, poets, and philosophers that shaped his style.[3] For in his temperament and competitive youth, perhaps eager to pursue the fame and success that humanism fostered, he prided himself on his skills as a forger. He borrowed ideas from drawings and sculptures, imitating and modifying them, not for fraudulent purposes, but as a demonstration of his *virtù*—his technical skills, analytical intelligence, and his command of various styles.[4] As observed from his biographies, letters, and sonnets, Michelangelo was also an avid reader of vernacular literature and a dedicated learner of classical philosophy, absorbing much of Dante's, Petrarch's, and Plato's words and synthesizing theirs with biblical teachings to create his artistic and literary masterpieces.[5] In this context, it can be said that because the young Michelangelo

[1] Giorgio Vasari, *The Lives of the Artists* (New York: Oxford University Press, 1991), 473.
[2] Fred S. Kleiner and Helen Gardner, "Renaissance and Mannerism in Cinquecento Italy," in *Gardner's Art Through the Ages: A Global History*, 14th ed. (Boston: Wadsworth Publisher, 2014), 499.
[3] Ascanio Condivi, *The Life of Michelangelo* (University Park: Pennsylvania State University Press, 1999), 11.
[4] Condivi, *The Life of Michelangelo*, 10.
[5] Vasari, *The Lives of the Artists*, 474.

dedicated himself to studying his predecessors' and contemporaries' sculptures and paintings, in addition to attending to the subjects and figures whose philosophies provided the foundation for or a relationship with such works, the older and more experienced Michelangelo could then form his unique aesthetic philosophy.[6]

While much has been written about his paintings and sculptures, in addition to the letters and sonnets he had authored throughout the course of his life, Michelangelo rarely commented on his masterpieces and the artistic processes behind them. To further understand Michelangelo's aesthetic philosophy, it is important to explore the artist's values and motivations that drove him to create his art. He rejected the kind of grace and detachment espoused by Baldassarre Castiglione (1478-1529) and Raphael Santi (1483-1520), instead cultivating an image of himself as hardworking, intense, and passionate.[7] His desire to reveal—rather than conceal through the act of *sprezzatura* ("effortless grace")—the labor, agony, and ecstasy behind his art can be seen in works completed in his early career prior to 1505 such as the *Bacchus* (1496-97, Bargello, Florence), the *St. Peter's Pietà* (1498–1500, Bargello, Florence), and the *David* (1501–1504, Accademia Gallery, Florence).[8]

By examining Giorgio Vasari's *Lives of the Most Excellent Painters, Sculptors, and Architects* (1550, 1568) and Ascanio Condivi's *Life of Michelangelo* (1553), I will explore how Michelangelo's background formed his psychological understanding of himself, and how they fueled his passion to search and present divine beauty in his early Roman and Florentine sculptures. I will also examine the letters and sonnets from the high Roman period in praise of Tommaso dei Cavalieri's (1509-1587) physical beauty and Vittoria Colonna's (1492-1547) spiritual nobility to further explain Michelangelo's belief in the Neoplatonic idea of pressing from outward beauty—*il bel del fuor che agli occhi piace* ("the outward beauty that is pleasing to the eyes")—to reveal the hidden abstract form of beauty—*trascenda nella forma universal* ("transcendent in the universal form").[9] Through understanding the material and spiritual beauty that Michelangelo sought to present in his works, we can gain a deeper understanding of his aesthetic philosophy, and how it promotes art as a domain of religious experience and a medium to strive for divine perfection.

2. Michelangelo's Background and Early Roman and Florentine Sculptures

To explore the values and motivations behind Michelangelo's creative process, it is necessary to mention his family background and social aspirations. Michelangelo claimed that he belonged to a noble lineage—one that had some eminence in Florence but had descended in the social scale over time. Many members of the Buonarroti family, who were then widely believed to be the descendants of the counts of Canossa, had been *signori* ("lords"), or as Condivi explained, "members of the supreme magistracy of the [Florentine] republic."[10] Michelangelo's father,

[6] Condivi, *The Life of Michelangelo*, 102.
[7] Walter Pater, *"The Poetry of Michelangelo"* (New York: Garland Publishing, 1999), 520-530.
[8] Frederick Hartt and David G. Wilkins, "The Cinquecento," in *History of Italian Renaissance Art*, 7th ed. (Upper Saddle River: Prentice Hall, 2011), 469.
[9] Pater, *"The Poetry of Michelangelo,"* 535.
[10] Condivi, *The Life of Michelangelo*, 6.

Ludovico di Leonardo Buonarroti Simoni (1444-1534), was also the *podestà* ("chief magistrate") of Chiusi and Caprese in the Casentino valley, although his appointment only lasted for a year.[11] In Italy and throughout European civilization at the time, a family's status established an individual's status. Therefore, it was not unreasonable that Michelangelo's father and uncles resisted his artistic inclinations; a profession in the arts was not fitting for one of aristocratic lineage, and the young boy should have aspired for a more elevated occupation.[12] They finally relented and ceased their harsh judgements when Michelangelo's skills caught the eye of Lorenzo de' Medici (1449–1492), the "Magnificent," then the *de facto* ruler of the Florentine republic and a great patron of the arts, who took the young Michelangelo under his wing for two years.[13]

To the ambitious and proud Michelangelo, it was important for him to recover the family's lost status and wealth, whether real or imagined, and he made it his life-long preoccupation to do so with his artistic genius. Michelangelo was well aware of his worth and talents, and when this confidence was coupled with being born a Buonarroti and having connections to other noble Florentines since his father was a cousin of Lorenzo, it formed a strong and autonomous character. He was not only admired, but feared for his temper, for he spared neither high nor low emotions. He dealt with potential patrons more or less as equals, and a remarkable proportion of his youthful sculptures was produced on his own initiative.[14] He lived on a few commissions, although this was obtained partly by the help of his family, friendships, and patronage ties.[15] The Medicis were especially important to Michelangelo's early career, providing him with commissions and opportunities that permitted the young Michelangelo to pursue an unconventional path independent of the guild system and the highly competitive artisan profession.[16] Due to his familial lineage and unique opportunities that set him apart from most of his fellow artists, Michelangelo was particularly sensitive about being treated like an artisan despite being one himself.

In 1548, Michelangelo wrote a letter to Lionardo Buonarroti, his nephew and heir, asking him to dismiss an acquaintance for having addressed a letter to him as "the Sculptor Michelangelo." "Tell him," he asked of Lionardo, "not to address his letters to the sculptor Michelangelo, for here I am known only as Michelangelo Buonarroti ... I have never been a painter or sculptor, in the sense of having kept a shop ... although I have served the popes; but this I did under compulsion."[17] The pride Michelangelo had towards his family's illustrious reputation and his own independence was best shown in his works. Michelangelo did not provide the products and services typical of Renaissance artists throughout his career, such as painted chests, *cassone*, for the home or artworks for establishments like the guilds. Even in his youth, he produced a series of unique objects that were considered innovative and difficult to imitate, such as the *Bacchus* (figure 2) and the *St. Peter's*

[11] Condivi, *The Life of Michelangelo*, 6.
[12] Vasari, *The Lives of the Artists*, 420.
[13] Condivi, *The Life of Michelangelo*, 12-13.
[14] Condivi, *The Life of Michelangelo*, 12-13.
[15] Vasari, *The Lives of the Artists*, 421.
[16] Vasari, *The Lives of the Artists*, 421.
[17] Michelangelo, and George Bull, *Life, Letters, and Poetry* (New York: Oxford University Press, 2009).

Pietà (figure 3).[18] As Michelangelo did not conform to the general expectations that high society had towards artisans and their work at the time, Michelangelo's aesthetic philosophy was not bound by any set of traditions and rules. Change was also the primary constant that appeared in his masterpieces, which correlated to his changing values and perspective of the world. Thus, one can see clear differences between the *Bacchus* (figure 2) and the *St. Peter's Pietà* (figure 3), which are different from one another in content, style, and effect.

Together with the *St. Peter's Pietà* (figure 3), the *Bacchus* (figure 2) is one of only two surviving sculptures from the artist's first period in Rome, and it is in this monumental figure that we see Michelangelo's first attempt on his quest towards discovering divine beauty. According to Vasari, when Cardinal Raffaele Riario, the Cardinal of San Giorgio (1460-1521), commissioned this sculpture, Michelangelo understood that his work was destined to be a part of an open-air collection of antiquities, where subjects were designed to be viewed in the round.[19] Paying homage to the antique sculptures in whose company it would stand, Michelangelo granted the *Bacchus* attributes that would have accompanied a similar figure by a Greco-Roman artist.[20] Thus, he provided vegetal attributes such as grapes, vine leaves, and ivy as ornaments for the *Bacchus*'s crown, and in the god's right hand, a cup that refers to the intoxicating liquor obtained from the grapes.

However, Michelangelo did not content himself solely with creating a sculpture that looked like a product of antiquity, and indeed, by the standards of its day, the *Bacchus* is an unusual sculpture in every respect. Instead of lending the figure the characteristic feature of authentic large-scale statues from classical antiquity—a firm *contrapposto* ("counterpoise")—the *Bacchus* provides a different stance: with the god's inebriation and unsteady gait in the insecure position of his legs and in the angle of his torso and head, the statue seems to make a mockery of classical *contrapposto*.[21] A hundred years of Renaissance sculptural tradition was suddenly taken to an extreme by Michelangelo's creativity. In adherence to the usual Renaissance preference and his patron's expectations, Michelangelo lent the mystery of antique cults to his *Bacchus*. Yet, he questioned the ideals of ancient sculpture through the exaggerated *contrapposto* and the unconventional portrayal of the divine figure.[22] While Michelangelo's patron Cardinal Raffaele Riario and his contemporaries found the product of his imagination and its divergence from general expectations to be appalling or perhaps borderline offensive, the *Bacchus* nevertheless sheds light on the heart of Michelangelo's aesthetic philosophy: that the figure's striking yet unflattering portrayal—a realistic depiction of the effects of intoxication fitting for the god of drunkenness and excess—could be understood in a metaphorical sense as a pathway to the revelation of divine mysteries.[23]

Although Michelangelo's name was already spreading throughout Rome when he dealt with lighter subject matter in sculpting the *Bacchus*, his fame grew when he produced an unmistakable emblem of Catholic devotion, the *St. Peter's Pietà* (figure 3). Michelangelo was barely twenty-four years' old when the *Pietà* was completed and installed in the Basilica di San Pietro, therefore it was not surprising

---

[18] William Wallace, *The Genius of Michelangelo* (Virginia: The Teaching Company, 2007), Lecture 24.
[19] Vasari, *The Lives of the Artists*, 421.
[20] William Wallace, *The Genius of Michelangelo* (Virginia: The Teaching Company, 2007), Lecture 7.
[21] Wallace, *The Genius of Michelangelo,* Lecture 7.
[22] Wallace, *The Genius of Michelangelo,* Lecture 7.
[23] Wallace, *The Genius of Michelangelo,* Lecture 7.

that it cemented Michelangelo's fame as the leading sculptor of his era and earned him the nickname "Il Divino" ("Divine One").[24] Despite the hard material of marble he had to work with, Michelangelo challenged himself to imitate the soft layers, folds, and smallest undulations of the Virgin's draperies in all of their tactile wealth, in addition to the detailed muscles, veins, and nerves stretched over the framework of the bones of the dying Christ.[25] At the time, the *Pietà* was a common subject depicted throughout Northern Europe, and as its compositional type originated in the religious devotion and mysticism of the Middle Ages, there was an expectation that these works should be expressive to directly convey emotional and physical pain.[26] Many of the Northern European renderings of the *Pietà* were awkward, stiff, with the lifeless body of Christ precariously placed on the lap of his grieving mother. However, just as he took inspiration from Classical Antiquity yet diverged from it to sculpt the *Bacchus*, Michelangelo took the compositional type of the Northern conventions to sculpt the *Pietà* and translated it into an aesthetically different form.[27] Instead of the crude realism sculptors from the Middle Ages employed to craft their versions of the *Pietà*, Michelangelo conveyed the typical expression of suffering and grief through an artistically fashioned beauty in adherence to Renaissance ideals.

For Michelangelo, the art of antiquity was not just an ideal to be admired, but also a touchstone and challenge; he wanted his works to allude to ancient customs yet at the same time surpass them with novel elements. This is evidenced alone by Michelangelo's representation of the Virgin which in addition to being technically perfect is distinguished by her particularly youthful and placid features despite her age. When asked about why he chose to depart from the norm, Michelangelo replied:

> Don't you know that women who are chaste remain much fresher than those who are not? How much more so a virgin who was never touched by even the slightest lascivious desire which might alter her body? Indeed, I will go further and say that this freshness and flowering of youth, apart from being preserved in her in this natural way, may also conceivably have been given divine assistance in order to prove to the world the virginity and perpetual purity of the mother ... Therefore you should not be surprised if, with this in mind, I made the Holy Virgin, mother of God, considerably younger in comparison with her Son than her age would ordinarily require ...[28]

Thus, it can be said that while crafting the *Bacchus* and the *St. Peter's Pietà*, Michelangelo was not only concerned with how he portrayed the two sculptures' external beauty, but also how their aesthetic form could shed light on the message he wanted to convey. By taking inspiration from ancient traditions and building upon them with his imagination, Michelangelo provided examples of how external beauty could reflect inward beauty—one that lies underneath the façade and mirrors

[24] Wallace, *The Genius of Michelangelo*, Lecture 7.
[25] Vasari, *The Lives of the Artists*, 425.
[26] Wallace, *The Genius of Michelangelo*, Lecture 7.
[27] Wallace, *The Genius of Michelangelo*, Lecture 7.
[28] Condivi, *The Life of Michelangelo*, 27.

divine perfection. In addition, despite the clear formal and thematic differences that distinguish the two works, their variety introduces other constants that provided foundation for his aesthetic philosophy: Michelangelo's search for originality, his poetic sensibility, coupled with his desire to extend the expressive possibilities of sculpture that reference other fields of the arts and humanities.

Although Michelangelo's genius as a sculptor had already been proven two years earlier when he completed the *St. Peter's Pietà*, it was the completion of the *David* (figure 4) that cemented his fame and success. The political symbolism of Michelangelo's work had been present from the start; the commission for the sculpture and the choice that it would be placed in the Duomo, a building of the highest civic and religious importance, were connected to the civil and political upheavals Florence had experienced a few years ago.[29]

Like the *Bacchus* and the *St. Peter's Pietà*, Michelangelo was not content with repeating the usual conventions previous Renaissance masters and contemporaries have employed while crafting the *David*, and indeed, the sculpture was an unusual product of its day. Unlike the frail and effete shepherd boy depicted by Andrea del Verrocchio (1435-1488) or Donatello (1386-1466), Michelangelo's biblical *David* is a grown and muscular young man, somewhat akin to the figures carved by the Greco-Roman ancients.[30] Instead of portraying the moment directly after David's victory over Goliath, Michelangelo decided to depict the moment before—a detail that is also suggested by his furrowed brow and his left hand loosely holding the slack and empty sling.[31]

With its monumentality, its accentuated muscular frame, and its *contrapposto*, Michelangelo's *David* recalls the antique representations of Hercules and Apollo and thereby the cardinal virtue of *fortitudo* associated with such heroes and deities. According to Katie Kressner, due to the "rediscovery of ancient Greco-Roman sculptures, there was a mania to make artistic bodies contain every perfection and every sublimity."[32] However, it was not just the "taut, muscular, and poised Olympian perfection" that Michelangelo wanted to imitate, but also the "heavenly perfection of Christ, who was at once Man and became God that he wanted to reveal."[33] Michelangelo fusing a biblical subject with a Greco-Roman representation demonstrates that in his mind, David's perfection as the prototype of a "Christian" ruler was only assured by his combination of strength and beauty together. This could perhaps mean that when Michelangelo carved the *David*, he intended for the sculpture to not only serve as a symbol of Florentine pride but also of all humanity raised to a new power: a godlike grandeur and beauty that does not ignore but rather emphasizes the faults that come with being human.[34] The idea behind this sculpture resonates with the Neoplatonic ideals that Michelangelo believed in: that despite the imperfections that humans are born with, they make the most of their destinies as God has given them the right to do so, and to reveal the beauty and potential that lies within humans' spirits is to form a direct connection with God. Michelangelo's sculpture, with its simultaneously powerful and aesthetically perfect physique, sheds light on Michelangelo's aesthetic

[29] Vasari, *The Lives of the Artists*, 427.
[30] Hartt and Wilkins, *History of Italian Renaissance Art*, 477.
[31] Hartt and Wilkins, *History of Italian Renaissance Art*, 477.
[32] Katie Kresser, "Bodies, Beauty and Time: On Michelangelo," Christian Scholar's Review, February 8, 2021, https://christianscholars.com/bodies-beauty-and-time-on-michelangelo/. (accessed August 20, 2022).
[33] Kressner, "Bodies, Beauty and Time: On Michelangelo."
[34] Vasari, *The Lives of the Artists*, 427.

philosophy: that art is nothing but a reflection of divine perfection, and it is the artist's mission to link physical beauty with divine destiny.

3. Michelangelo's Relationships with Tommaso dei Cavalieri and Vittoria Colonna

To further explain Michelangelo's aesthetic philosophy and the Neoplatonic idea of pressing from outward beauty, *il bel del fuor che agli occhi piace* ("the outward beauty that is pleasing to the eyes"), to reveal the hidden abstract form of beauty, *trascenda nella forma universal* ("transcendent in the universal form"), it is important to consider the letters and sonnets that he authored from the high Roman period in praise of Cavalieri's physical beauty and Vittoria Colonna's spiritual nobility.

Over the course of his life, Michelangelo had written more than 400 poems, which revealed his inner turmoil and the complicated emotions that fueled his creative process. Many of these letters and sonnets were directed to Tommaso dei Cavalieri (figure 5), who was described as one of Michelangelo's "most beautiful young men, his dearest and most honest friends."[35] Cavalieri was often associated with two characteristics—beauty and *virtù*—which gave him a reputation that he was proud of. He used it to form connections with great artists and thinkers. In his first letter to Michelangelo, Cavalieri said: "I do believe, nay I am certain, that the cause of the affection you have for me is this: that since you are most virtuous–or better, an embodiment of *virtù* itself–you are compelled to love those who believe in it, and love it, including myself ..."[36] According to Donato Gianotti, Michelangelo also shared Cavalieri's attitude by stating "Every time I see someone who has some *virtù* ... I see myself compelled to love him and do so in abandonment, so that I am no longer myself, but all his."[37]

Michelangelo's interactions with Cavalieri left an enduring mark on him.[38] However, Michelangelo's preference for male company and male beauty was treated obliquely by his biographers such as Condivi and Vasari as they were writing in the wake of the religious fervor sparked by the Protestant Reformation (1517-1648) and the Council of Trent (1545-63).[39] He was a devout Catholic and was nearing sixty when he fell for a young nobleman forty years his junior, and while he managed to combine his love of God and love of Cavalieri's beauty, the fact he directed various intimate letters and private drawings to Cavalieri raised many questions among society.[40] However, despite High Renaissance society raising many questions about Michelangelo's alleged homosexuality, Condivi once defended Michelangelo's love for the male body as to be in the platonic realm, and philosophically and artistically motivated, rather than romantic:

> He [Michelangelo] has also loved the beauty of the human body as one who knows it extremely well, and loved it in such a way as to

[35] Carmen Bambach, "*The Poetry of Michelangelo*" (New York: The Metropolitan Museum of Art, 2017), 136.
[36] Marcella Marongiu, "*Tommaso de' Cavalieri*" (New York: The Metropolitan Museum of Art, 2017), 287.
[37] Marcella Marongiu, "*Tommaso de' Cavalieri,*" 287.
[38] George Bull, *Life, Letters, and Poetry*, 142.
[39] Carmen Bambach, *"The Poetry of Michelangelo,"* 135.
[40] Carmen Bambach, *"The Poetry of Michelangelo,"* 136.

> inspire certain carnal men, who are incapable of understanding the love of beauty except as lascivious and indecent, to think and speak ill of him. It is as though Alcibiades, a very beautiful young man, had not been most chastely loved by Socrates, of whom he was wont to say that, when he lay down with him, he arose from his side as from the side of his father. I have often heard Michelangelo converse and discourse on the subject of love and have later heard from those who were present that what he said about love was no different than what we read in the writings of Plato.[41]

Condivi then elaborated further on Michelangelo's love for beauty:

> ... he has loved not only human beauty but everything beautiful in general: a beautiful horse, a beautiful dog, a beautiful landscape, a beautiful plant, a beautiful mountain, a beautiful forest, and every place and thing which is beautiful and rare of its kind, admiring them all with marveling love and selecting beauty from nature as the bees gather honey from flowers, to use it later in his works.[42]

Therefore, Michelangelo's poetry and gift drawings for Cavalieri show that the ardor of his sentiments are expressed simply in the purity of Neoplatonic terms, and that his love and celebration of the beauty of the male nude body is a crucial factor of his aesthetic philosophy and his understanding of the divine.

Although Cavalieri admired Michelangelo's otherworldly genius and valued his friendship with the artist tremendously, Cavalieri was heterosexual, had a family of his own, and based on the letters, most likely did not reciprocate the same passionate intensity in the relationship like Michelangelo had towards him. It can be said that this unrequited love followed Michelangelo for the rest of his life, but in his later years Michelangelo had abandoned any hopes of earthly satisfaction from this attachment by embracing Counter Reformation austerity and becoming a more devout Christian.[43] Perhaps, Michelangelo may have felt guilty about his emotions towards his understanding of the relationship between him and Cavalieri, recognizing with regret that his passion and desire for Cavalieri was, in spite of the Neoplatonic justifications, largely a physical, homoerotic, and unreciprocated desire—which was not the way to reach salvation. As Michelangelo's heart craved beauty just as much as he craved the divine—and to let go of his earthly attachments to Cavalieri and resolve his conflicted soul—he struck up a friendship and became closer to Vittoria Colonna, Marquess of Pescara (figure 5).

Born in 1492 into an old Roman family, Colonna was the widow of Ferrante Francesco d'Avalos (1490-1525), the scion of one of the oldest noble families of Italy. Condivi once spoke of their intimate relationship and poetic correspondence in *The Life of Michelangelo*:

> In particular, he greatly loved the marchioness of Pescara, whose sublime spirit he was in love with, and she returned his love

[41] Carmen Bambach, *"The Poetry of Michelangelo,"* 135.
[42] Condivi, *The Life of Michelangelo*, 105.
[43] Vasari, *The Lives of the Artists*, 478.

> passionately. He still has many of her letters, filled with honest and most sweet love, and these letters sprang from her heart, just as he also wrote many many sonnets to her, full of intelligence and sweet desire. She often traveled to Rome from Viterbo and other places where she had gone for recreation and to spend the summer, prompted by no other reason than to see Michelangelo; and he in return bore her so much love that I remember hearing him say that his only regret was that, when he went to see her as she was departing this life, he did not kiss her forehead or her face as he kissed her hand.[44]

Unlike Cavalieri, who dazzled Michelangelo with his beauty and had a rapturous effect on him, the traits that fueled Michelangelo's attraction and love to the noblewoman was not her appearance, but her spiritual goodness. As a celebrated poet and a devout Christian, Colonna had reshaped and imbued in Michelangelo a number of influences resonant in his time. She influenced his ideas about Neoplatonism, the language of the arts and literature, and several religious reform doctrines. She patronized his work, serving as one of his closest friends and his only female confidante. The relationship with Colonna, an accomplished woman of the High Renaissance and an acclaimed spiritual poet, even spurred Michelangelo to write some of his most inspirational poems.

With this evidence, one can begin to see the differences between Michelangelo's relationship with Colonna as opposed to the one with Cavalieri. While both relationships were characterized by a long-standing friendship and based upon mutual admiration and gifts, it can be seen that Colonna's role in transforming Michelangelo's spiritual and aesthetic philosophy was far more influential compared to Cavalieri's role. The letters Michelangelo and Colonna exchanged, in addition to the nature of their relationship, demonstrate that Michelangelo's understanding of himself, relationship to God, and aesthetic philosophy had changed more drastically.

Through his experiences with Cavalieri and Colonna, we can see that Michelangelo's aesthetic philosophy is based on one of the central tenets of Neoplatonic philosophy: that the two types of beauty that exist in relation to humans, one is physical and the other spiritual, vary in levels of importance. The physical beauty that Cavalieri possessed that so enraptured Michelangelo was just temporary, and would fade and disappear as time progressed. The spiritual beauty and *virtù* that Michelangelo tried to seek in Cavalieri and found in Colonna was eternal, and only could be found when one looks past all of the outward characteristics and looks into the soul of the person, thus looking at the divinity of God. Thus, if the soul is beautiful, that person will remain beautiful regardless of what happens to the rest of their body, and this was also demonstrated through the young face of Mary in Michelangelo's *St. Peter's Pietà*. However, it would be false to completely disregard the importance of physical beauty, as to according to Michelangelo and another central tenet of Neoplatonism, the body is the outward expression of the soul, partaking of its beauties and mirroring its passions, in addition to being a reflection of the divine. This was what Michelangelo was concerned with when he was carving the *David*–something that he also discussed with Colonna via one of his letters.[45]

---

[44] Condivi, *The Life of Michelangelo*, 103.
[45] George Bull, *Life, Letters, and Poetry*, 200.

In adherence to the aesthetic and philosophical culture of Neoplatonism, Michelangelo's works and letters communicated the artist's intense emotions and located these feelings within a broader system of values that interlinks passion, beauty, and the divine. Michelangelo saw art and beauty differently than many artists of the High Renaissance; to have obtained a deep sense of humanistic beauty and appreciation to envision the body, its gesture, and its sense of movement just by looking at a block of marble is not a skill that every person can attain. In this way, he followed the footsteps of the Greco-Romans before him, who were known for their fascination with the nude human form and finding great beauty within it. Michelangelo, however, not only alluded to the ancient customs, but surpassed them. He not only infused his figures with unrivaled energy, but also saw in them spiritual power and opportunities to uncover elements of the divine. Michelangelo's aesthetic philosophy then becomes the epitome of Neoplatonic ideology: it is not only occupied with beauty and human potential, but also promotes art as a domain of religious experience and a medium to strive for divine perfection. Through this, Michelangelo's aesthetic philosophy espouses the idea that we, as humans, are the embodiment of possibility, and that while we are shaped by the environment we live in, we are nevertheless above nature due to our divine intellectual abilities and can shape it in return.

Appendix

Figure 1. Il Passignano (Dom-enico Crespi), *Portrait of Mich-elangelo*, Early 17th century, Oil on canvas, Collection of Galleria Enrico Lumina, Bergamo.

Figure 2. Michelangelo, *Bacchus and Pan*, 1496, Sculpture, Museo del Bargello, Florence, Italy.

Figure 3. Michelangelo, *St. Peter's Pietà*, 1497–1499, Sculpture, Basilica di San Pietro in Vaticano, Rome.

Figure 4. Michelangelo, *David*, 1501–1504, Sculpture, Accademia Gallery in Firenze, Florence.

Figure 5. Michelangelo or Daniele da Volterra, *Portrait of Tommaso dei Cavalieri*, Black chalk drawing, Musée Bonnat-Helleu, Bayonne.

Figure 6. Sebastiano del Piombo, *Portrait of Vittoria Colonna*, 1520-1525, Oil on wood, Rome.

Works Cited

Bambach, Carmen. "Private Works: the Master of 'Disegno' and his Gift Drawings." In *Michelangelo: Divine Draftsman and Designer*, edited by Carmen Bambach, 130-68. New York, NY: The Metropolitan Museum of Art, 2017.

Condivi, Ascanio. *The Life of Michelangelo*. Edited by Hellmut Wohl. Translated by Alice Sedgewick Wohl. University Park, PA: Pennsylvania State University Press, 1999.

Hartt, Frederick, and David G. Wilkins. "The Cinquecento." Essay. In *History of Italian Renaissance Art*, 7th ed., 477. Upper Saddle River: Prentice Hall,2011.

Kleiner, Fred S., and Helen Gardner. "Renaissance and Mannerism in Cinquecento Italy." Essay. In *Gardner's Art Through the Ages: A Global History*, 14th ed., 499. Boston, MA: Wadsworth Publisher, 2014.

Kresser, Katie. "Bodies, Beauty and Time: On Michelangelo." Christian Scholar's Review, February 8, 2021. https://christianscholars.com/bodies-beauty-and-time-on-michelangelo/. (accessed August 20, 2022).

Marongiu, Marcella. "Tommaso de' Cavalieri." In *Michelangelo: Divine Draftsman and Designer*, edited by Carmen Bambach, 287-89. New York, NY: The Metropolitan Museum of Art, 2017.

Michelangelo, and George Bull. *Life, Letters, and Poetry*. Translated by George Bull and Peter Porter. New York, NY: Oxford University Press, 2009.

Vasari, Giorgio. *The Lives of the Artists.* Translated by Julia Conaway Bondanella and Peter Bondanella. New York, NY: Oxford University Press, 1991.

Pater, Walter. "The Poetry of Michelangelo." In *Michelangelo: Selected Readings*, edited by William Wallace, 523-42. New York, NY: Garland Publishing, 1999.

Wallace, William E. *The Genius of Michelangelo*, Lecture 7. Chantilly, Virginia: The Teaching Company, 2007.

Wallace, William E. *The Genius of Michelangelo*, Lecture 24. Chantilly, Virginia: The Teaching Company, 2007.

A Proposal to Implement Cas-CLOVER Technology in the Treatment of Patients with Spinal Muscular Atrophy

Alp Namalan

Author Background: *Alp Namalan grew up in Turkey and currently attends Galatasaray High School, in Istanbul, Turkey. His Pioneer research concentration was in the field of biology and titled "Cell Biology."*

Abstract

Spinal Muscular Atrophy (SMA) is an autosomal recessive neuromuscular disease associated with insufficient survival motor neuron (SMN) protein levels caused by the deletion of the SMN1 gene. The physical symptoms of SMA include muscular weakness and severe impairment of motor functions. Currently, existing therapies aim to prevent further complications instead of correcting the underlying genetic cause of SMA. This proposed study, on the contrary, investigates the novel Cas-CLOVER genome-editing technology as a treatment method that addresses the root cause of the disease without causing off-target mutations. To achieve that, the Cas-CLOVER system will be cloned into plasmid vectors to be injected into mouse zygotes. The predicted results indicate elevated SMN protein levels, prolonged lifespan, and improved motor functions in treated mouse pups compared to untreated SMA mice. Furthermore, the off-target cutting rate for Cas-CLOVER is expected to be insignificant in contrast to other genome-editing tools. The findings of this proposed study theoretically prove the efficacy of Cas-CLOVER in the treatment of patients with SMA and related diseases.

1. Introduction

Spinal Muscular Atrophy (SMA) is a rare motor neuron disease mainly affecting children. It is one of the most common autosomal recessive diseases and the primary genetic cause of infant mortality. One out of every 10,000 babies is affected by the illness, and one out of every 50 persons is a carrier. This incidence rate is almost twice as high in developing countries, such as Turkey. Caused by the homozygous deletion, or mutation, of the survival motor neuron 1 (SMN1) gene, SMA is characterized by motor neuron death in the spinal cord,

resulting in gradual muscular weakening and, in severe instances, respiratory failure and death. The disease is divided into numerous categories based on severity and age of onset. SMA Type I is the most prevalent and gravest form: babies show symptoms as soon as they are born, never learn to sit, and usually do not live past the age of two. Other types have a later onset and are less severe. SMA has no current cure, and treatment can only be used to manage symptoms (Tisdale & Pellizzoni, 2015).

The most recent breakthroughs in SMA therapy include the substitution of the SMN1 gene, modification of SMN2 splicing (a homologous gene that mostly encodes for non-functional proteins), and upregulation of muscle growth. Currently, the most effective treatment for SMA is gene therapy (Zolgensma), which involves the introduction of a healthy SMN1 gene in patients (Schorling et al., 2020). Although it has successfully prolonged the survival period and improved motor functions in patients, Zolgensma has serious limitations. It cannot reverse the damage to motor neurons, carries a serious risk of hepatotoxicity, and is priced at a staggering $2.125 million for a one-dose injection (Chand et al., 2020; Starner et al., 2019).

Genome editing might show promise as a viable treatment for the condition as prenatal SMA screening is becoming more common. CRISPR/Cas9 has previously been proven to promote SMN expression in human induced pluripotent stem cells (iPSCs) and extend the survival period of mice (Li et al., 2020). Due to its high risk of causing off-target mutations, however, CRISPR/Cas9 is not the perfect candidate for SMA treatment. Recently, another genome-editing tool, Cas-CLOVER, has been developed. Studies using this tool on human T-cells demonstrated that Cas-CLOVER had comparable efficiency to CRISPR/Cas9 while eliminating off-target alterations (Li et al., 2019). This paper aims to propose a mechanism that implements the Cas-CLOVER technology to treat SMA patients in the prenatal stage. Editing the SMN2 gene could elevate the levels of functional SMN proteins without requiring regular injections, resulting in a healthy phenotype. Additionally, the absence of off-target aberrations would minimize adverse effects and lead to a better safety profile.

2. Overview of Spinal Muscular Atrophy

2.1. Epidemiology of SMA

SMA is a rare genetic neuromuscular condition that involves a mutation in the survival motor neuron gene (SMN1) in chromosome 5q. The disease causes alpha motor neuron death, which leads to increasing muscular weakness. In approximately 92% of the patients, homozygous deletion of SMN1 is responsible for the disorder (Alías et al., 2009). SMA is subdivided into different categories based on the age of onset and maximum motor achievement possible. Infants with Type I SMA (Werndig-Hoffman disease) exhibit symptoms by six months, are never able to sit, and cannot survive past 2 years due to respiratory complications. Patients with Type II form can sit or even stand, but never attain the capacity to walk. Type III (Kugelberg-Welander disease) patients have a later onset and usually survive until adulthood. Type IV is a rare adult-onset form

that causes mild motor dysfunction. (Finkel et al., 2015). The incidence of SMA is 1:6000 to 1:10,000 live births. Although Type I is the most common form, around half the clinically registered patients are affected by Type II SMA. This might be due to the extremely short life expectancy of Type I patients (Verhaart et al., 2017).

2.2. Genetic Basis of SMA

Most SMA cases are caused by the homozygous deletion of exon 7 in the 5q13.2 region of chromosome 5. This leads to insufficient production of the SMN protein, which is essential for motor neuron maintenance (Baker et al., 2019). Only a homologous gene, SMN2, is responsible for SMN production in SMA patients. However, due to a single C to T transition in this gene, exon 7 is usually skipped in SMN2 splicing, which causes a truncated SMN protein to form. The mutation interrupts an exon splice enhancer sequence. This results in the formation of an exonic splicing silencer that binds to the splicing repressor heterogeneous ribonuclear protein (hnRNP) A1. Nevertheless, approximately 10% of proteins encoded by SMN2 are functional. An individual possesses zero to eight copies of the SMN2 gene, and the copy number of SMN2 is inversely related to disease severity in SMA patients (Farrar & Kiernan, 2015).

2.3. Molecular Mechanisms in SMA

2.3.1. Regulation of SMN2 Expression

One way to raise SMN production to normal levels in SMA patients is to upregulate SMN2 expression. Trans-acting factors, such as ELK-1, CREB, and STAT5 play a crucial role in this regulation. Inhibition of the overactivated MEK/ERK/ELK-1 pathway has been proven to prevent motor neuron death and improve the phenotype of mice models (Branchu et al., 2013). Activation of the AKT/CREB pathway by insulin-like growth factor-1 receptor (Igf-1r) also resulted in increased neuroprotection (Biondi et al., 2015). In another study, the Janus kinase (JAK)/STAT pathway was activated with prolactin (PRL), which increased SMN production (Farooq et al., 2011).

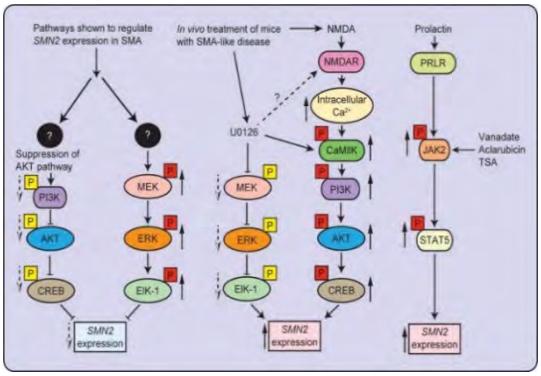


Figure 1. *Pathways in the regulation of SMN2 gene expression (Ahmad et al., 2016).*

2.3.2. Neurodegeneration Caused by Low Levels of SMN

Several cellular pathways lead to SMA pathogenesis resulting from low SMN levels. One of them is the RhoA (a GTPase molecule)/Rho-associated protein kinase (ROCK) pathway, which is essential for cytoskeletal regulation. Low SMN levels cause ROCK to be activated, altering the cytoskeletal organization and resulting in neurodegeneration (Bowerman et al., 2007). The c-Jun NH2-terminal kinase (JNK) pathway also plays a role in neurodegeneration by disrupting microtubule stability and causing axonal defects. Genetic repression of this molecule rescued phenotypes in mice with SMA (Genabai et al., 2015). Additionally, the interaction between SMN and the mRNA-binding protein HuD unravels the cellular mechanism of SMA pathogenesis. It promotes the trafficking and localization of poly(A)-mRNA along the axon. A mutation in the Tudor domain of SMN disrupts the interaction, which results in motor neuron death (Fallini et al., 2011).

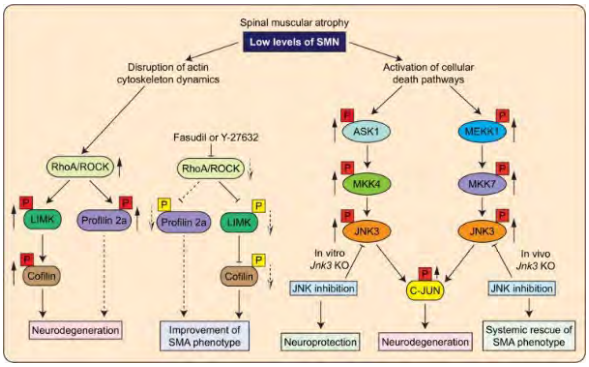


Figure 2. *Neurodegeneration mechanisms in SMA (Ahmad et al., 2016).*

2.3.3. Modifier Proteins

In addition to the molecular pathways mentioned above, studies have identified two modifier proteins that might play key roles in SMA pathogenesis and influence disease severity. The first modifier protein, plastin 3 (PL3), located on Xq23, raises F-actin levels and promotes axonogenesis. Its expression also increases during neuronal differentiation (Oprea et al., 2008).

The second modifier, zinc finger protein 1 (ZPR1), located in the 11q23.3 region, is a protein that interacts with SMN. SMN is found in the cytoplasm and subnuclear bodies, such as gems and Cajal bodies. It is also essential for the biogenesis of spliceosomal small nuclear ribonucleoproteins (snRNPs). ZPR1 is necessary, along with SMN, for the localization of nuclear bodies. It also forms complexes with SMN and snRNPs, which are involved in pre-mRNA splicing. SMA patients have lower ZPR1 expression, resulting in motor neuron death (Gangwani et al., 2001).

The study of the molecular mechanisms and modifier proteins in SMA is instrumental in discerning the cellular consequences of the condition and developing new methods to supplement or replace current therapeutical approaches.

2.4. Medical Complications Caused by SMA

In people with SMA, the loss of motor neurons in the spinal cord causes hypotonia and muscle weakness. Symptoms usually depend on disease type. Patients with SMA Type I have severe hypotonia, symmetrical flaccid paralysis, and no head control. They cannot sit without assistance. Paradoxical breathing and a bell-shaped upper torso occur from the sparing diaphragm paired with reduced intercostal muscles. Moreover, bulbar denervation causes tongue fasciculation and weakness, as well as difficulty sucking and swallowing. It also reduces airway protection and raises the risk of aspiration pneumonia. Type II patients attain the ability to sit. With assistance, a few can stand, but none can walk on their own. Patients suffer from severe scoliosis, which requires medical intervention. Fine tremors, accompanied by finger extension or gripping, are also typical. Poor swallowing might affect weight gain. As with Type I patients, removing tracheal secretions and coughing may become difficult because of poor bulbar function and weak intercostal muscles. Respiratory insufficiency is a common cause of mortality during adolescence. Patients with SMA Type III exhibit a wide range of symptoms. They usually reach all significant motor milestones, such as walking independently. Scoliosis is also common in these patients. Type III patients show symptoms of joint overuse, which is usually accompanied by weakness. Lastly, SMA Type IV has an onset during adulthood and patients only show mild motor dysfunction (Lunn & Wang, 2008).

3. Current Therapies for SMA Patients

3.1. Modification of SMN2 Splicing

The FDA has approved two drugs that aim to modify SMN2 splicing: Nusinersen (Spinraza) and Risdiplam (Evrysdi). Nusinersen is an antisense oligonucleotide that binds to a particular region in the intron, downstream of exon 7, on the SMN2 pre-mRNA. This changes the SMN2 mRNA transcript's splicing to include exon 7, resulting in more full-length SMN proteins produced. A study found that exon 7 is included in 15–26% of the SMN2 transcripts in thoracic spinal cord tissue from untreated newborns with SMA or infants with no illness. On the contrary, exon 7 is included in 50–69% of SMN2 transcripts from babies with SMA who had been exposed to nusinersen (Finkel et al., 2016). Nusinersen is administered intrathecally, which results in post-lumbar puncture syndrome as a common adverse effect (Cordts et al., 2020). As the drug requires regular administration, this might affect the quality of life in patients.

Risdiplam is another FDA-approved SMN2 splicing modifier that binds on the exonic splicing enhancer 2 (ESE2) and 5′ splice site (5'ss) locations within the exon 7. Binding to the 5'ss improves U1 snRNA binding. The interaction with ESE2 is thought to cause the hnRNP G to dislocate, allowing the U1 snRNP complex to bind. As a result, full-length proteins that include exon 7 are produced. A study involving 21 infants with SMA showed that risdiplam use improved motor functions and increased the life span of patients. The drug requires daily oral administration and might cause fever, diarrhea, and rash (Markati et al., 2022). It might also lead to off-target complications.

Additionally, both drugs are highly costly for patients ($125,000 per dose for nusinersen and $340,000 per year for risdiplam), considering that they require regular administration (Chaytow et al., 2021).

3.2. Replacement of SMN1

The other FDA-approved drug in SMA treatment, onasemnogene abeparvovec (Zolgensma), is a form of gene therapy. It involves a single intravenous injection of self-complementary adeno-associated virus (scAAV9) vectors carrying a healthy copy of the SMN1 gene. As homozygous deletion of the SMN1 gene causes SMA, reintroduction of this gene in SMA patients shows promise. In a study with 15 SMA patients, all of the patients receiving gene therapy outlasted the 20-month milestone. Only 8% of patients with the condition would generally survive past 20 months without permanent ventilation (Mendell et al., 2017). Zolgensma is approved for patients younger than 2 years of age. It has a price of $2,125,000 for a single injection, which might make it extremely difficult for patients to access gene therapy (Chaytow et al., 2021). Moreover, Zolgensma is known to cause serious hepatotoxicity by elevating aminotransferase levels. A study found that an additional intake of prednisolone (a corticosteroid) with a dosage of 1 mg/kg/day might alleviate these effects (Chand et al., 2021). However, the administration of corticosteroids in gene therapy might also result in adverse health outcomes.

3.3. SMN-Independent Therapies

SMN is a component of the machinery that assembles spliceosomal components. Thus, its deficiency causes a general splicing deficit, with motoneurons being particularly vulnerable. Splicing, on the other hand, is not limited to neurons or motoneurons. It is plausible that the impairment of the SMN protein's "housekeeping" job, more specifically, its participation in the neuronal actin cytoskeleton, might impact all cells and organs. As drugs approved for treatment may be insufficient for reversing the damage caused by SMA, several treatment methods that target SMA-specific disruptions downstream of SMN deficiency might supplement or replace SMN-related therapies (Hensel et al., 2020). Along with non-specific treatments that elevate SMN levels, molecules that promote neuroprotection or target muscles, neuromuscular junctions (NMJs), the cytoskeleton, or cell death pathways are in development. Although most of these molecules have not been tested or resulted in significant improvements, combining them with SMN-related therapies might improve disease phenotype in patients (Chaytow et al., 2021). For example, as histone acetylation is an essential epigenetic factor that influences SMN expression, histone deacetylase (HDAC) inhibitors are examined in SMA models. A study showed that the HDAC inhibitor LBH589 could repair the improper splicing of SMN2 and upregulate SMN expression, especially when administered with suboptimal doses of Nusinersen (Pagliarini et al., 2020).

4. Background on Cas-CLOVER Genome Editing

4.1. Genome Editing for SMA Treatment

The CRISPR-Cas9 system uses an RNA-guided nuclease to edit the human genome. Single-guide RNAs (sgRNAs) direct the Cas9 (or Cpf1) endonucleases to bind a specified genomic region near a protospacer neighboring motif. This results in a double-strand break (DSB). Non-homologous end joining (NHEJ) or homology-driven repair can then be used to repair the DSB (HDR) (Long et al., 2016). At this point, researchers have examined CRISPR/Cas9 in the treatment of numerous diseases, including SMA. A study involving 36 mouse zygotes showed that 20 of the 36 surviving SMA pups born had NHEJ alterations at the SMN2 gene. 17 of these pups were rescued for SMA clinical signs and lived for more than 100 days. Furthermore, the splicing-corrected mice had a much longer median lifetime (>400 days) than the unedited control SMA mice (which was only 13 days). The same study also found that CRISPR/Cas9 prevents the degeneration of motor neurons derived from human stem cells (Li et al., 2020).

4.2. Efficacy of Cas-CLOVER over CRISPR/Cas9

When implementing the CRISPR/Cas9 system for biomedical and therapeutic applications, off-target mutations are detected more often than the desired mutation (≥50%). This may induce genomic instability and alter the activity of otherwise normal genes (Zhang et al., 2015). The Cas-CLOVER system uses a dual gRNA-guided nuclease, in which each half-site subunit contains a fusion protein of a catalytically inactive Cas9 (dCas9) and the restriction endonuclease Clo051, instead of a single guide RNA (gRNA) for sequence-specific guidance of CRISPR/Cas9 binding and cutting. Clo051 activity is dependent on the formation of a dimer, and thus, DNA cleavage requires simultaneous on-target binding of two different gRNA-guided endonucleases within a specified proximity. In a study investigating the effect of Cas-CLOVER on T cells, next-generation sequencing (NGS) revealed no off-target mutations (Li et al., 2019).

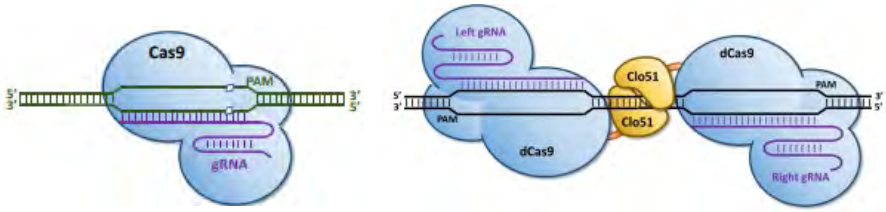


Figure 3. *Comparison of CRISPR/Cas9 (left) and Cas-CLOVER (right) (Li et al., 2019).*

4.3. Limitations of Cas-CLOVER

Although Cas-CLOVER has proven efficient in changing the genome, it may still have limitations worth considering. The leading concern about Cas-CLOVER is

that only a few studies have investigated this genome-editing tool. Other studies might disprove the efficacy of Cas-CLOVER. Thus, it is not possible to know whether, compared to CRISPR/Cas9, Cas-CLOVER will have similar effects on increasing SMN levels. Secondly, despite its higher specificity, Cas-CLOVER might not be exempt from other limitations that are associated with CRISPR/Cas9, such as DNA-damage toxicity, immunotoxicity, and difficulties in delivery (Uddin et al., 2020). Moreover, the exorbitant price of current genome-editing therapies is a significant concern. A study involving liver-focused genome-editing therapy found that treatment could cost up to $1.8 million (Wilson & Carroll, 2019). However, it should be noted that genome-editing technologies are still evolving and will probably become more affordable in the near future.

5. Methodology

5.1. Mouse Model Selection

Animal models are essential resources for investigating the molecular biology and neuropathology of SMA and for pre-clinical evaluation of treatment approaches. Since the neural systems of mice and humans are fairly similar in structure and function, mice have been widely acknowledged as useful models for SMA. Furthermore, the mouse Smn gene is a homolog of the human SMN1 gene (Bebee et al., 2012). Heterozygous moderate Type II (Smn+/-; SMN2+/+; SMNdelta7+/+, stock number: 005025) mouse pairs will be purchased from the Jackson Laboratory (Bar Harbor, ME, USA). The pairs will be bred to obtain mice with homozygous knockout of the Smn gene (Smn-/-; SMN2+/+; SMNdelta7+/+). Wild-type offspring (Smn+/+; SMN2+/+; SMNdelta7+/+) from the breeding pair will be used as controls. All mice will be exposed to a 12/12-h light/dark cycle.

5.2. Production of dCas9-Clo051 System

5.2.1. Guide RNA (gRNA) Design

The dCas9-Clo051 requires a dual gRNA system. In a study involving SMA mice, researchers designed gRNAs specific to intronic splicing silencer-N1(ISS-N1) and ISS+100 regions (Li et al., 2020). Guide RNA pairs will be generated based on the sequences below.

Table 1. *gRNA sequences designed to target ISS-N1 and ISS+100 regions (Li et al., 2020).*

Primer sequences (5' to 3')	Target
Forward-caccGAAGATTCACTTTCATAATGC Reverse-aaacGCATTATGAAAGTGAATCTTC	ISS-N1
Forward-caccGTCAGATGTTAGAAAGTTGAA Reverse-aaacTTCAACTTTCTAACATCTGAC	ISS+100

5.2.2. Plasmid Design

The plasmid vector for the experimental setup should express U6 promoter-driven gRNA and the fusion protein dCas9-Clo051. To that end, designed gRNA sequences will be cloned into a plasmid vector containing a U6 promoter (Addgene, #48962). A plasmid that contains the Clo051 nuclease combined with dCas9 will also be obtained. As Clo051 is not a widely studied protein, a plasmid vector with Clo051 currently remains hypothetical. However, given the fact that the FokI nuclease has numerous common characteristics with Clo051 (e.g. dimerization required for DNA cleavage), the vector is expected to be similar to FokI-dCas9 (Addgene, #52970) in structure (Bitinaite et al., 1998). Sequences from the two plasmids will be subcloned into a single final vector. The vector will then be amplified via polymerase chain reaction (PCR).

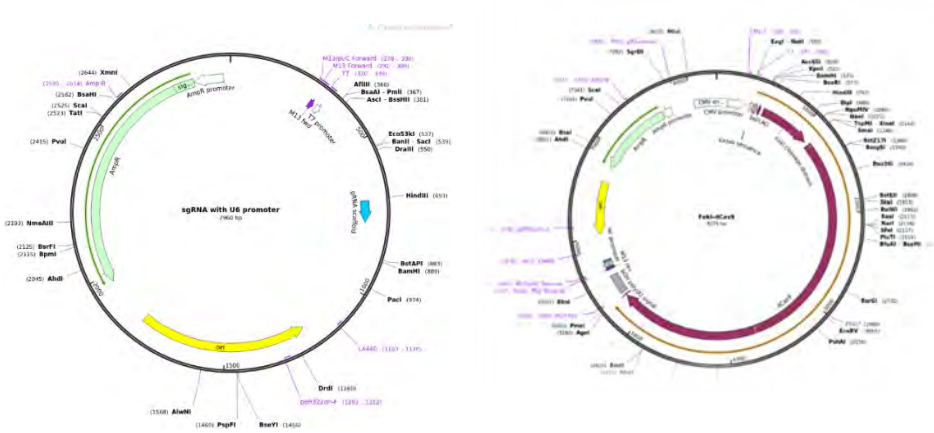


Figure 4. *Graphical maps of sgRNA with U6 promoter (Addgene, #48962) (left) and FokI-dCas9 (Addgene, #52970) (right) plasmid vectors.*

5.3. Zygote Injection and Transplantation

Protocols used by a previous study on gene editing in mice will be followed for the injection of plasmids into zygotes (Fujihara & Ikawa, 2014). Heterozygous female moderate Type II mice will be super-ovulated and mated to heterozygous male moderate Type II mice. Fertilized cells will be collected and cultured in a KSOM medium before injection. The pU6-gRNA-dCas9-Clo051-expressing plasmids will be diluted to 5 ng/μL and injected into the pronuclei of collected cells. KSOM medium at 37°C and 5% CO2 will be used to cultivate surviving eggs in vitro into two-cell stage embryos for 20 hours, which will then be transplanted into the oviduct of 0.5 d.p.c. pseudo-pregnant females.

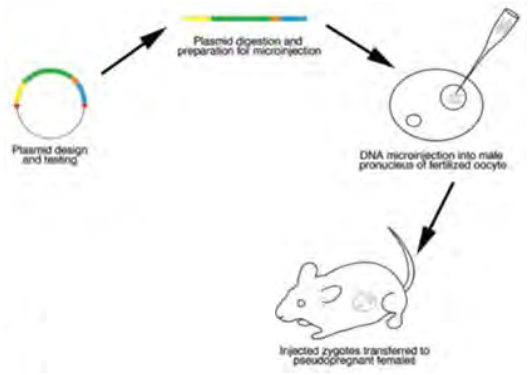


Figure 5. *Main steps in the generation of transgenic mice (Vacaru et al., 2014).*

6. Data to Be Collected

6.1. Genotyping Analysis

To determine whether the target gene has successfully been edited, genotyping analysis of mice is imperative. Therefore, genomic DNA will be extracted from the tail tips of 2-3 week-old mice by phenol/chloroform purification. PCR will be performed, and amplified fragments will be examined by gel electrophoresis. The products will then be sub-cloned into a plasmid vector and sequenced using Sanger sequencing. From each cell line, clones will be chosen, evaluated, and employed to determine the editing ratio.

6.2. Western Blot

Western blotting protocols will be performed as previously described (Zhou et al., 2018). Tissue samples from mice will be isolated and lysed with a radioimmunoprecipitation assay (RIPA). 1× phenylmethanesulfonyl fluoride (PMSF, 1mM) and peptidase inhibitor will be added to the RIPA buffer. The proteins obtained will be separated via 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and moved onto a polyvinylidene difluoride (PDVF) membrane. The membranes will undergo an overnight incubation at 4°C in an SMN-antibody solution, containing mouse-anti-SMN and anti-β-tubulin. The proteins will finally be visualized with an enhanced chemiluminescence (ECL) western blotting detection kit (Thermo Fisher Scientific).

6.3. Behavioral Tests

Reflex tests are effective predictors of typical development and can be used to evaluate the degree of neuronal maturation in developing mice. Therefore, the extent of motor neuron damage caused by SMA can be assessed through such tests (Fox, 1965). The righting reflex, clasping response, and grip strength tests

will be carried out as previously explained (Butchbach et al., 2007). For the righting reflex test, each pup will be placed on its back, and the amount of time it takes for all four paws to firmly touch the ground will be measured. The righting reflex latency will be measured every day, from post-natal day (PND) 2 until PND8. For the clasping test, each pup will be softly caressed on the forelimb and hindlimb footpads with a toothpick while being held by the scruff of its neck. It will be noted if a clasping reaction occurs or not. From PND2 through PND8, clasping reactions will be recorded. To gauge grip strength in mice, a suspension test will be used from PND 11 to PND14. Each pup will be put on a wire mesh, followed by an inversion of the mesh. It will be noted how long it takes the pup to loosen the mesh.

6.4. Off-Target Analysis

Off-target cleavage is one of the primary concerns in genome editing. To examine the off-target mutations, genomic DNA will be collected from mice. The potential off-target locations of the designed gRNAs will be determined based on their off-target scores according to the Zhang Lab website (http://crispr.mit.edu), as previously described (Anderson et al., 2018). Top off-target sequences will be selected and the determined sequences will then be amplified by multiplex PCR. Amplified sequences from mutation-positive mice will be screened by targeted amplicon next-generation sequencing (NGS).

7. Predicted Results

7.1. Statistical Analysis

All data are displayed as mean ± standard deviations (SD). One-way analysis of variance (ANOVA) was used to evaluate the statistical significance of the difference in data obtained from several groups. P-values were then computed and only those less than 0.05 were considered statistically significant.

7.2. Disruption of the Intronic Splicing Silencers

To correct the splicing of the SMN2 gene, two gRNAs will be designed to disrupt the intronic splicing silencers in intron 7: ISS-N1 and ISS+100 (Fig. 6A). The gRNAs will be cloned into a plasmid vector, along with dCas9 and Clo051, and injected into the zygotes collected from heterozygous moderate Type II mice. Genomic DNA isolated from 2-3 week-old mice will be sequenced to determine whether SMN2-ISSs have successfully been disrupted. As shown in Figure 6B, approximately one-third of all examined SMN sites are predicted to contain an altered ISS. Nevertheless, this number is adequate for the correction of SMN2 splicing.

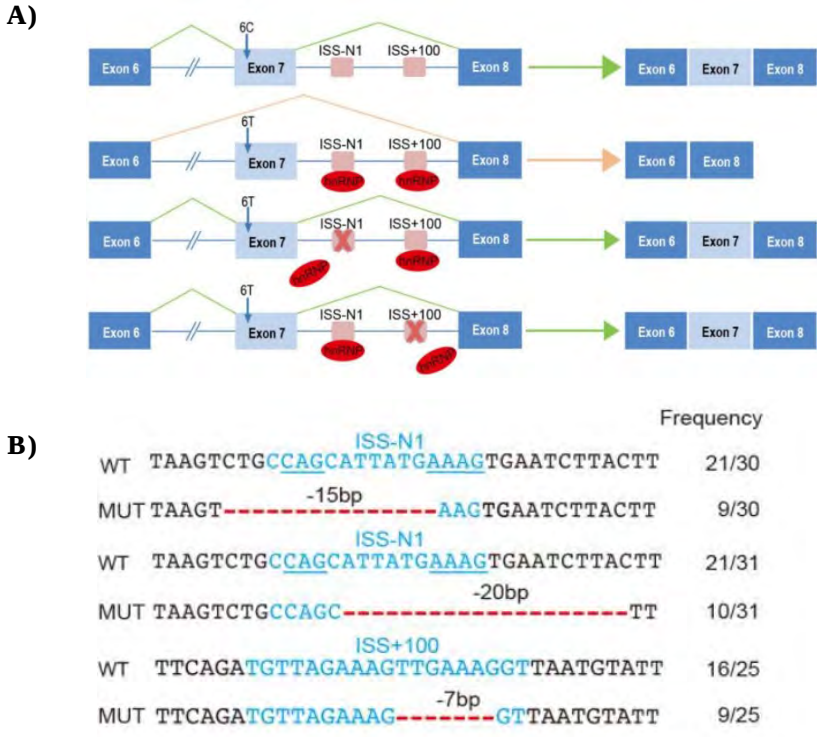


Figure 6. *Correction of exon 7 splicing via disruption of ISS-N1 and ISS+100 in SMA mice. (A) Diagram demonstrating the disruption of ISS-N1 and ISS+100 in intron 7 of the SMN 2 gene (SMN1, SMN2, gRNA 1, and gRNA 2 from top to bottom). (B) Sequence alignments from SMA mouse cells with ISS-N1 and ISS+100 disruptions. The red dashed lines denote the deletions, whereas the blue lines denote the ISS-N1 core motif sequences. The percentage of the pertinent genotype is shown in the column on the right (Li et al., 2020).*

7.3. SMN Expression

Western blot analysis will be used to assess SMN expression in splicing-corrected mice at PND9. In contrast to SMA mice, all of the splicing-corrected mice are expected to have higher exon 7 inclusion rates (~50% vs. ~5%) and immunoblotting analysis should verify the elevated SMN protein accumulation (Fig. 7). A significant variance in the SMN levels across different treated mouse lines is anticipated. Genotyping analysis shows, however, that this difference might be attributed to variations in the edited sequences (Li et al., 2020).

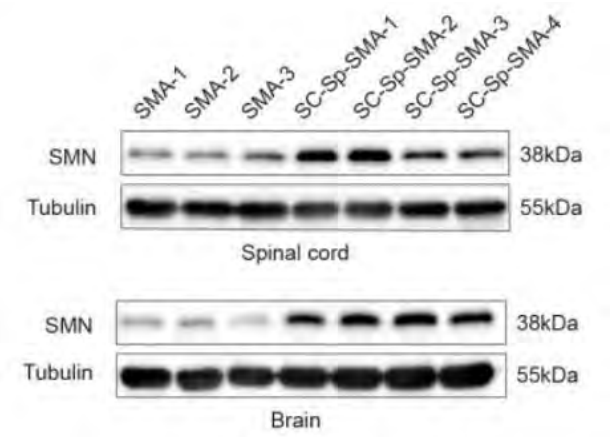


Figure 7. *Immunoblotting of the SMN protein extracted from spinal cord and brain tissues of mice. The first three columns represent samples from SMA mice, whereas the other four columns represent those from splicing-corrected (SC-Sp-SMA) mice (Li et al., 2020).*

7.4. Prevention of SMA

7.4.1. Survival Rate and Lifespan

Of all injected SMA embryos, ~80% of them are predicted to result in a live birth, compared to ~85% of wild-type embryos. Starting the first post-natal day, the body weight and lifespan of live-born mice will be assessed daily. In contrast to wild-type mice that usually live past 400 days, untreated SMA mice had the shortest survival period (~15 days). The splicing-corrected mice, on the other hand, are expected to have a substantial increase in their median lifespan, reaching > 170 days (Fig. 8A). Although genome editing should considerably improve the lifespan, the treated mice might still have a shorter survival period compared to their wild-type counterparts (Rashnonejad at al., 2019).

7.4.2. Body Weight

Since SMA mice cannot survive past 15 days, the body weight of all groups will be compared at PND 15. It is hypothesized that, at PND 15, wild-type controls and treated mice should have a considerably higher body weight than SMA mice (5.75 ± 0.75 g; 4.3 ± 0.7 g; and 2.2 ± 0.3 g, respectively) ($p < 0.001$). While the body weight of the treated mice is predicted to be moderately lower than the control group over time, it should still be closer to that of healthy mice than untreated SMA mice (Fig. 8B) (Rashnonejad et al., 2019).

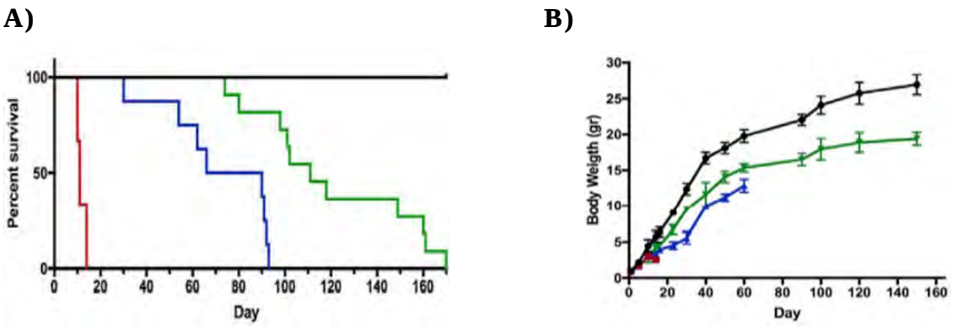


Figure 8. *Percent survival and body weight of SMA mice. (A) Survival curves for wild-type (black), SMA (red), and splicing-corrected mice (blue and green). (B) Body weight of wild-type (black), SMA (red), and splicing-corrected mice (blue and green) (Rashnonejad et al., 2019).*

7.4.3. Motor Function

To evaluate the motor function of the treated mice, the righting reflex, clasping response, and grip strength tests will be performed. For the righting reflex test, it is predicted that the splicing-corrected mice should recover their position faster than their untreated SMA counterparts. The splicing-corrected mice are also expected to exhibit improved grip strength in contrast to SMA mice. However, since the clasping response test will be carried out from PND2 to PND8, prior to the onset of motor neuron death, the splicing-corrected and SMA mice should demonstrate no significant differences in tactile sensory behavior, such as clasping of the forepaw or hind paw after moderate stimulation (Butchbach et al., 2007).

7.5. Off-Target Mutations

Given that off-target activity is an essential issue in genome editing, potential off-target locations of the gRNAs will be examined for indel mutations by PCR amplification and next-generation sequencing (NGS). Considering the structure of Cas-CLOVER that requires dimerization before DNA cleavage, it is hypothesized that the off-target mutation rate should be substantially lower than that of CRISPR/Cas9. The predicted mean indel frequency at examined locations should either range from ~ 0.01% to ~ 0.09% or not be at a statistically significant rate at all (Fig. 9).

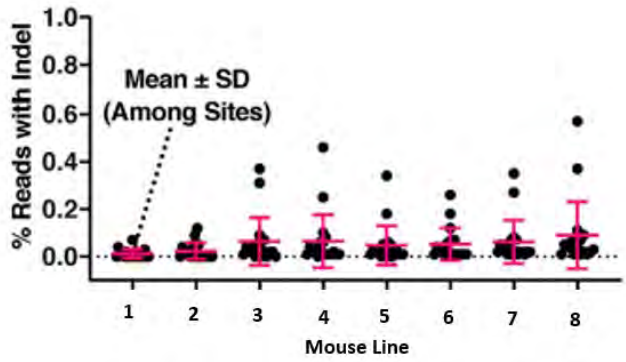


Figure 9. *Rates of off-target activity across different treated mouse lines, displayed with the mean percentage of reads with indel (Madison et al., 2022).*

8. Discussion

8.1. Principle of the Proposed Study

This research proposal aims to define the implementation of a novel genome-editing technology as a cure for SMA. To date, the therapeutical approaches designed to treat SMA (e.g. nusinersen and onasemnogene abeparvovec) have been limited to the treatment of physical symptoms caused by the syndrome without the correction of underlying genetic origins (Gyngell et al., 2020). Another major concern is the inadequacy of current therapies in completely restoring the motor neuron damage inflicted by SMA, which highlights the need for novel methods that target the genetic origins of this disease before the onset of motor neuron damage. Rendered possible by emerging screening technologies, genome editing at the embryonic stage is a promising method that could potentially cure SMA. Currently, CRISPR/Cas9 is the most widely used genome editing tool. Several studies have already proved that it has been successful in increasing SMN protein levels and, therefore, rescuing the SMA phenotype. Nevertheless, the fact that CRISPR/Cas9 generates almost as many genetic aberrancies as it corrects remains a major drawback. Given its specific design that eliminates the risk of off-target mutations, Cas-CLOVER, a novel genome-editing tool, should be superior to previous therapeutical approaches in targeting the underlying genetic origins of SMA without resulting in unforeseen alterations in the genome.

8.2. Interpretation of Predicted Results

The predicted results of this in vivo experiment demonstrated that ISS-disrupted SMA mice exhibited increased survival period, body weight, and motor function. As the genotyping analysis showed that only one in every three SMN loci underwent ISS disruption, the Clo051-mediated genome editing method might initially appear ineffective. However, it should be noted that SMA

symptoms stem from impaired RNA splicing mechanisms associated with inadequate levels of the SMN protein. Therefore, examining protein expression levels would result in a more accurate assessment of the efficacy of the proposed method. The western blotting analysis showed that Cas-CLOVER led to a ten-fold increase in SMN levels, thereby proving the viability of the novel genome-editing tool. The splicing-corrected mice had comparable lifespan and body weight to healthy mice and were almost identical to their wild-type counterparts in terms of motor functioning. Furthermore, the treatment of mice with Cas-CLOVER merely resulted in a trivial off-target mutation rate. These findings suggest that Cas-CLOVER would successfully rescue the SMA phenotype with high fidelity.

Compared to previous studies involving the treatment of SMA mice, this experiment yielded comparable results in the improvement of SMN protein levels, lifespan, and motor functions. A study that investigated the effect of scAAV9-mediated gene therapy on SMA treatment demonstrated that treated mice had a prolonged median survival of 199 days and increased SMN protein levels in the brain tissue (~40% of wild-type mice) (Dominguez et al., 2011). Similarly, in this proposed study, treated SMA mice had a median lifespan of >170 days and SMN protein levels reaching ~50% of healthy mice. Furthermore, both studies showed that the treatment of SMA mice resulted in a substantial enhancement of motor functioning. As far as off-target activity, however, Cas-CLOVER had a vast advantage over other genome-editing tools, which is attributable to its design that requires on-target dimerization before DNA cleavage. In this experiment, the off-target mutation rate of Cas-CLOVER did not exceed ~0.1%, which is significantly lower than that of CRISPR/Cas9 (~13%) (Madison et al., 2022). These reinforce the hypothesis that Cas-CLOVER would be a promising tool to cure SMA without causing inconveniences associated with current drugs (e.g. physical side effects) or off-target cutting of the genome, which might result in more damage than the treatment restores.

8.3. Limitations and Future Implications

The most evident concern regarding Cas-CLOVER technology is the lack of studies delving into the implementation of this novel technology in different areas. To date, only the effect of Cas-CLOVER on T cells has been studied (Madison et al., 2022). This might raise doubts about the efficacy of this novel genome-editing technology on other types of cells. Since Cas-CLOVER is not identical to CRISPR/Cas9 in terms of the functioning mechanism, it might not be suitable to correct the genetic mutations that give rise to SMA. Another major limitation is that, currently, systemic genome editing of human zygotes is nearly impossible due to technical and ethical concerns. Genome editing might also lead to unexpected adverse effects in humans that have not been recorded in mice. Finally, the R&D costs of a novel, Cas-CLOVER-mediated SMA therapy may render the treatment extremely expensive. Based on the standard R&D costs of a drug, this proposed treatment method might initially cost more than a million dollars per patient. Due to the lack of studies involving Cas-CLOVER in the current literature, future studies should mainly focus on new experiments to investigate the effect of Cas-CLOVER on various types of cells as well as its potential to treat different diseases.

9. Conclusion

Given the current potential of novel genome-editing technologies to treat various genetic disorders and the predicted results of this proposed study, Cas-CLOVER-mediated therapy shows promise as an effective method to prevent the SMA phenotype in newborns. In addition to its fidelity, Cas-CLOVER appears to provide a more fundamental solution to SMA in comparison to existing therapies, targeting the root cause of the disease: Cas-CLOVER addresses the genetic origins of SMA, unlike current drugs that focus on managing the physical symptoms of the condition. Despite the ethical concerns surrounding genome editing and the necessity of future studies to further prove its efficacy, Cas-CLOVER is likely to become a viable treatment method for SMA patients in the near future.

References

Ahmad, S., Bhatia, K., Kannan, A., & Gangwani, L. (2016). Molecular mechanisms of neurodegeneration in spinal muscular atrophy. *Journal of experimental neuroscience*, 10, JEN-S33122.

Alías, L., Bernal, S., Fuentes-Prior, P., Barceló, M. J., Also, E., Martínez-Hernández, R., ... & Tizzano, E. F. (2009). Mutation update of spinal muscular atrophy in Spain: molecular characterization of 745 unrelated patients and identification of four novel mutations in the SMN1 gene. *Human genetics*, 125(1), 29-39.

Anderson, K. R., Haeussler, M., Watanabe, C., Janakiraman, V., Lund, J., Modrusan, Z., ... & Warming, S. (2018). CRISPR off-target analysis in genetically engineered rats and mice. *Nature methods*, 15(7), 512-514.

Baker, M. et al. Maximizing the Benefit of Life-Saving Treatments for Pompe Disease, Spinal Muscular Atrophy, and Duchenne Muscular Dystrophy Through Newborn Screening: Essential Steps. *JAMA Neurol.* (2019). doi:10.1001/jamaneurol.2019.1206

Bebee, T. W., Dominguez, C. E., & Chandler, D. S. (2012). Mouse models of SMA: tools for disease characterization and therapeutic development. *Human genetics*, *131*(8), 1277-1293.

Biondi, O., Branchu, J., Salah, A. B., Houdebine, L., Bertin, L., Chali, F., ... & Charbonnier, F. (2015). IGF-1R reduction triggers neuroprotective signaling pathways in spinal muscular atrophy mice. *Journal of Neuroscience*, 35(34), 12063-12079.

Bitinaite, J., Wah, D. A., Aggarwal, A. K., & Schildkraut, I. (1998). Fok I dimerization is required for DNA cleavage. Proceedings of the national academy of sciences, 95(18), 10570-10575.

Bowerman, M., Shafey, D., & Kothary, R. (2007). Smn depletion alters profilin II expression and leads to upregulation of the RhoA/ROCK pathway and defects in neuronal integrity. *Journal of molecular neuroscience*, 32(2), 120-131.

Branchu, J., Biondi, O., Chali, F., Collin, T., Leroy, F., Mamchaoui, K., ... & Charbonnier, F. (2013). Shift from extracellular signal-regulated kinase to AKT/cAMP response element-binding protein pathway increases survival-motor-neuron expression in spinal-muscular-atrophy-like mice and patient cells. *Journal of Neuroscience*, 33(10), 4280-4294.

Butchbach, M. E., Edwards, J. D., & Burghes, A. H. (2007). Abnormal motor phenotype in the SMNΔ7 mouse model of spinal muscular atrophy. *Neurobiology of disease*, 27(2), 207-219.

Chand, D., Mohr, F., McMillan, H., Tukov, F. F., Montgomery, K., Kleyn, A., ... & Kullak-Ublick, G. (2021). Hepatotoxicity following administration of onasemnogene abeparvovec (AVXS-101) for the treatment of spinal muscular atrophy. *Journal of Hepatology*, 74(3), 560-566.

Chaytow, H., Faller, K. M., Huang, Y. T., & Gillingwater, T. H. (2021). Spinal muscular atrophy: From approved therapies to future therapeutic targets for personalized medicine. *Cell Reports Medicine*, 2(7), 100346.

Cordts, I., Lingor, P., Friedrich, B., Pernpeintner, V., Zimmer, C., Deschauer, M., & Maegerlein, C. (2020). Intrathecal nusinersen administration in adult spinal muscular atrophy patients with complex spinal anatomy. *Therapeutic advances in neurological disorders*, 13, 175628641988 7616.

Dominguez, E., Marais, T., Chatauret, N., Benkhelifa-Ziyyat, S., Duque, S., Ravassard, P., ... & Barkats, M. (2011). Intravenous scAAV9 delivery of a codon-optimized SMN1 sequence rescues SMA mice. *Human molecular genetics*, 20(4), 681-693.

Fallini, C., Zhang, H., Su, Y., Silani, V., Singer, R. H., Rossoll, W., & Bassell, G. J. (2011). The survival of motor neuron (SMN) protein interacts with the mRNA-binding protein HuD and regulates localization of poly (A) mRNA in primary motor neuron axons. *Journal of Neuroscience*, 31(10), 3914-3925.

Farooq, F., Molina, F. A., Hadwen, J., MacKenzie, D., Witherspoon, L., Osmond, M., ... & MacKenzie, A. (2011). Prolactin increases SMN expression and survival in a mouse model of severe spinal muscular atrophy via the STAT5 pathway. *The Journal of clinical investigation*, 121(8), 3042-3050.

Farrar, M. A., & Kiernan, M. C. (2015). The genetics of spinal muscular atrophy: progress and challenges. *Neurotherapeutics*, 12(2), 290-302.

Finkel, R. S., Chiriboga, C. A., Vajsar, J., Day, J. W., Montes, J., De Vivo, D. C., ... & Bishop, K. M. (2016). Treatment of infantile-onset spinal muscular atrophy with nusinersen: a phase 2, open-label, dose-escalation study. *The Lancet*, 388(10063), 3017-3026.

Finkel, R., Bertini, E., Muntoni, F., & Mercuri, E. (2015). 209th ENMC international workshop: outcome measures and clinical trial readiness in spinal muscular atrophy 7–9 November 2014, Heemskerk, The Netherlands. *Neuromuscular Disorders*, 25(7), 593-602.

Fox, W. M. (1965). Reflex-ontogeny and behavioural development of the mouse. *Animal behaviour*, 13(2-3), 234-IN5.

Fujihara, Y., & Ikawa, M. (2014). CRISPR/Cas9-based genome editing in mice by single plasmid injection. In *Methods in enzymology* (Vol. 546, pp. 319-336). Academic Press.

Gangwani, L., Mikrut, M., Theroux, S., Sharma, M., & Davis, R. J. (2001). Spinal muscular atrophy disrupts the interaction of ZPR1 with the SMN protein. *Nature cell biology*, 3(4), 376-383.

Genabai, N. K., Ahmad, S., Zhang, Z., Jiang, X., Gabaldon, C. A., & Gangwani, L. (2015). Genetic inhibition of JNK3 ameliorates spinal muscular atrophy. *Human molecular genetics*, 24(24), 6986-7004.

Gyngell, C., Stark, Z., & Savulescu, J. (2020). Drugs, genes and screens: The ethics of preventing and treating spinal muscular atrophy. *Bioethics*, 34(5), 493-501.

Hensel, N., Kubinski, S., & Claus, P. (2020). The need for SMN-independent treatments of spinal muscular atrophy (SMA) to complement SMN-enhancing drugs. *Frontiers in neurology*, 11, 45.

Li, J. J., Lin, X., Tang, C., Lu, Y. Q., Hu, X., Zuo, E., ... & Chen, W. J. (2020). Disruption of splicing-regulatory elements using CRISPR/Cas9 to rescue spinal muscular atrophy in human iPSCs and mice. *National science review*, 7(1), 92-101.

Li, X., Wang, X., Tong, M., Tan, Y., Down, J. D., Shedlock, D. J., & Ostertag, E. M. (2019). Cas-CLOVER™: A high-fidelity genome editing system for safe and efficient modification of cells for immunotherapy. In 2018 Precision CRISPR Congress Poster Presentation, Boston, MA.

Long, C., Amoasii, L., Bassel-Duby, R., & Olson, E. N. (2016). Genome editing of monogenic neuromuscular diseases: a systematic review. *JAMA neurology*, 73(11), 1349-1355.

Lunn, M. R., & Wang, C. H. (2008). Spinal muscular atrophy. *The Lancet*, 371(9630), 2120-2133.

Madison, B. B., Patil, D., Richter, M., Li, X., Cranert, S., Wang, X., ... & Ostertag, E. M. (2022). Cas-CLOVER is a novel high-fidelity nuclease for safe and robust generation of TSCM-enriched allogeneic CAR-T cells. Molecular Therapy-Nucleic Acids.

Markati, T., Fisher, G., Ramdas, S., & Servais, L. (2022). Risdiplam: an investigational survival motor neuron 2 (SMN2) splicing modifier for spinal muscular atrophy (SMA). Expert Opinion on Investigational Drugs, 31(5), 451-461.

Mendell, J. R., Al-Zaidy, S., Shell, R., Arnold, W. D., Rodino-Klapac, L. R., Prior, T. W., ... & Kaspar, B. K. (2017). Single-dose gene-replacement therapy for spinal muscular atrophy. *New England Journal of Medicine*, 377(18), 1713-1722.

Oprea, G. E., Kröber, S., McWhorter, M. L., Rossoll, W., Müller, S., Krawczak, M., ... & Wirth, B. (2008). Plastin 3 is a protective modifier of autosomal recessive spinal muscular atrophy. *Science*, 320(5875), 524-527.

Pagliarini, V., Guerra, M., Di Rosa, V., Compagnucci, C., & Sette, C. (2020). Combined treatment with the histone deacetylase inhibitor LBH589 and a splice-switch antisense oligonucleotide enhances SMN2 splicing and SMN expression in spinal muscular atrophy cells. *Journal of Neurochemistry*, 153(2), 264-275.

Rashnonejad, A., Chermahini, G. A., Gündüz, C., Onay, H., Aykut, A., Durmaz, B., ... & Özkınay, F. (2019). Fetal gene therapy using a single injection of recombinant AAV9 rescued SMA phenotype in mice. *Molecular Therapy*, 27(12), 2123-2133.

Schorling, D. C., Pechmann, A., & Kirschner, J. (2020). Advances in treatment of spinal muscular atrophy–new phenotypes, new challenges, new implications for care. *Journal of neuromuscular diseases*, 7(1), 1-13.

Starner, C. I., & Gleason, P. P. (2019). Spinal muscular atrophy therapies: ICER grounds the price to value conversation in facts. *Journal of Managed Care & Specialty Pharmacy*, 25(12), 1306-1308.

Tisdale, S., & Pellizzoni, L. (2015). Disease mechanisms and therapeutic approaches in spinal muscular atrophy. *Journal of Neuroscience*, 35(23), 8691-8700.

Uddin, F., Rudin, C. M., & Sen, T. (2020). CRISPR gene therapy: applications, limitations, and implications for the future. *Frontiers in oncology*, 10, 1387.

Vacaru, A. M., Vitale, J., Nieves, J., & Baron, M. H. (2014). Generation of transgenic mouse fluorescent reporter lines for studying hematopoietic development. In *Mouse genetics* (pp. 289-312). Humana Press, New York, NY.

Verhaart, I. E., Robertson, A., Leary, R., McMacken, G., König, K., Kirschner, J., ... & Lochmüller, H. (2017). A multi-source approach to determine SMA incidence and research ready population. *Journal of neurology*, 264(7), 1465-1473.

Wilson, R. C., & Carroll, D. (2019). The daunting economics of therapeutic genome editing. *The CRISPR Journal*, 2(5), 280-284.

Zhang, X. H., Tee, L. Y., Wang, X. G., Huang, Q. S., & Yang, S. H. (2015). Off-target effects in CRISPR/Cas9-mediated genome engineering. *Molecular Therapy-Nucleic Acids*, 4, e264.

Zhou, M., Hu, Z., Qiu, L., Zhou, T., Feng, M., Hu, Q., ... & Liang, D. (2018). Seamless genetic conversion of SMN2 to SMN1 via CRISPR/Cpf1 and single-stranded oligodeoxynucleotides in spinal muscular atrophy patient-specific induced pluripotent stem cells. *Human gene therapy*, 29(11), 1252-1263.

Investigating the Mechanism by Which SARS-CoV-2 ORF3a Accessory Protein Mediates Lysosomal Exocytosis

Matthew Wang

Author Background: *Matthew Wang grew up in China and currently attends Shanghai High School International Division in Shanghai, China. His Pioneer research concentration was in the field of biology and titled "Pandemic: The New Coronavirus."*

Abstract

The SARS-CoV-2 virus is the causative agent of the COVID-19 pandemic. Although not necessary for viral replication, SARS-CoV-2 accessory proteins function in its pathogenesis. The ORF3a accessory protein is the largest accessory protein coded by the SARS-CoV-2 genome, it is a highly conserved domain in the subgenus *Sarbecovirus,* sharing a 72.7% homogeneity with SARS-CoV ORF3a. The primary function of ORF3a is to promote viral release through the lysosomal exocytosis pathway. Its structure as a viroporin allows it to act as an ion channel that is capable of modifying ion concentration to promote exocytosis. Furthermore, it has proven to interact with the HOPS complex to prevent lysosome maturation, and also the Ca2+ ion channel TRPML3 to promote viral release. In this proposed study, the mechanism by which ORF3a recruits ion channel TRPML3 is investigated. Simultaenously, the mutability of ORF3a in different *Sarbecoviruses* and amongst different variants is investigated based on its ability to inhibit the functioning of the VPS39 protein of the HOPS complex. This study further seeks to propose experiments that will elucidate the changes in SARS-CoV-2 transmissibility by investigating its ability to promote lysosomal exocytosis, and also the mechanism in which it does so.

1. Introduction and Background

Coronaviruses (CoVs) are positive-stranded RNA viruses that are taxonomically placed under the family Coronaviridae; the family is further divided into four genera: *Alpha-*, *Beta-*, *Delta-*, and *Gamma-coronavirus* (Kadam et al. 2021). In the past twenty years, there were two large-scale disease outbreaks caused by coronaviruses: the severe acute respiratory syndrome CoV (SARS-CoV) and the Middle East respiratory syndrome CoV (MERS-CoV), which took place in 2002 and 2012, respectively (Ciotti et al. 2019). The two pandemics were caused by viruses of zoonotic origin and were proven to be highly transmissible among human individuals, resulting in a wide toll of human lives. Notably, both belonged to the *Betacoronavirus* genus and targeted the human respiratory system (Kirtipal et

al. 2020).

More recently, in December 2019, a novel coronavirus was reported in Wuhan, China (Zhu et al. 2020). This novel coronavirus, named severe acute respiratory syndrome CoV-2 (SARS-CoV-2), was proven to be highly transmissible. The virus led to the emergence of a global pandemic, causing both economic and social stagnation. As of July 31st, 2022, there have been more than 570 million confirmed patients worldwide, out of which more than 6.3 million died due to the virus (World Health Organization, 2022). In response to this outbreak, researchers began to extensively investigate the SARS-CoV-2 genome, structure, and pathogenicity.

The SARS-CoV-2 genome includes six protein-encoding open reading frames (ORFs) that are shared by all coronaviruses. More notably, its genome also includes various unlabeled ORFs that are only present in *severe acute respiratory syndrome-related coronavirus* or the subgenus *Sarbecovirus* (Jungreis et al. 2021). This places SARS-CoV-2 and SARS-CoV viruses under the same species, SARS-CoV. These ORFs include five "accessory" proteins previously identified in other viruses of the species, namely, ORFs 3a, 6, 7a, 7b, and 8 (Wu et al. 2020). The entire SARS-CoV-2 genome encodes a total of 29 CoV-2 proteins. These include 16 non-structural proteins (NSP1-NSP16) located at the ORF1a and ORF1b region on the 5 'end of the genome, 4 structural proteins that are recognized in all coronaviruses, namely the spike (S), membrane (M), nucleocapsid (N), and envelope (E), and 9 accessory protein ORFs (3a, 3b, 6, 7a, 7b, 8, 9b, 9c, and 10) (Gordon et al. 2020).

However, the function of accessory proteins does not seem to be essential for viral replication. Instead, they play a crucial function in pathogenesis, more specifically in the evasion of the immune response (Redondo et al. 2021).

ORF3a is the largest accessory protein among the 9 accessory proteins coded by the SARS-CoV-2 genome. Genetically, the ORF3a has a nucleotide length of 825 base pairs (bp) and translates into a protein of 275 amino acids with a molecular weight of 31 kilodaltons (Zhang et al. 2022). Within the *Betacoronavirus* subgenus *Sarbecovirus*, SARS-CoV, and other related bat coronaviruses, ORF3a is highly conserved (Kern et al. 2021). SARS-CoV ORF3a, the most phylogenetically closely related protein, has a 72.7% homogeneity with the SARS-CoV-2 ORF3a (Issa et al. 2020). Structurally, the ORF3a protein is a viroporin, a type of integral membrane protein that acts as an ion channel, playing a possible role in promoting virus release (Bianchi et al. 2021). In addition to its function of acting as an ion channel, ORF3a is crucial for maximizing replication and virulence in both SARS-CoV-2 and SARS-CoV. The ORF3a protein possesses a wide range of highly conserved functional motifs that presumably contributes to its multi-functionalities such as ion channel activity, viral replication, and cytopathogenic effects that cause COVID-19 (Issa et al. 2020).

Most notably, the localization of SARS-CoV-2 ORF3a (referred to as ORF3a hereafter unless explicitly stated otherwise) on the plasma membrane (PM) and late-endosome/lysosome indicates that it could promote viral release through the lysosomal exocytosis pathway (Chen et al. 2021). The role of ORF3a in inducing viral exit can be mainly divided into two parts. The first is through hijacking autophagosomes and preventing their maturation; the second is through regulating intracellular conditions to promote lysosomal exocytosis. The former process helps the virus avoid the immune system, while the latter mediates the viral release from the host cell.

Normally, cellular autophagy is a method of surveillance that protects the

cell from pathogens and renews damaged organelles. Double-membrane autophagosomes engulf non-specific materials or selected cargos in the cytoplasm such as invading pathogens, damaged organelles, and protein aggregates (Lamb et al. 2013). After this, the fusion of autophagosomes with lysosomes induces the formation of degradative autolysosomes (Zhao and Zhang 2019). However, since SARS-CoV-2 viral exit is mediated through lysosomal exocytosis, it must ensure that the environment within lysosomes is not degradative. ORF3a targets the HOPS (Homotypic Fusion and Protein Sorting) complex component VPS39, preventing it from interaction with the autophagosome SNARE (Soluble N-ethylmaleimide-sensitive factor Attachment Protein Receptor) protein STX117(Antonin 2000). The HOPS complex has been proven to be responsible for mediating the autophagosome and lysosome fusion through interaction with STX17 (Jiang et al. 2014). The binding of ORF3a to the VPS39 competent of the HOPS complex thus prevents the formation of a trans-SNARE complex, indirectly disrupting the process of autophagosome-lysosome fusion (Miao et al. 2021). This is the process by which ORF3a prevents autophagosome maturation and prepares the lysosome environment for viral exit.
During the process of regular lysosomal exocytosis, BORC (BLOC (biogenesis of lysosome-related organelles complex)-one-related complex)-ARL8b complex helps transport lysosomes from the perinuclear regions to the PM (Wu et al. 2020) (Pu et al. 2016). Then, the SNARE complex mediates the fusion of lysosomes with the plasma membrane (PM). The SNARE complex includes the proteins VAMP7 (Vesicle Associated Membrane Protein 7), STX4 (Syntaxin 4), and SNAP23 (Synaptosomal-associated protein 23). These proteins together form a complex that serves as a binding site for the general membrane fusion machinery (Saftig and Klumperman 2009). The process of fusion would also require an increase in Ca2+ level, both intracellular and localized. ORF3a mediates lysosomal exocytosis by increasing the concentration of Ca2+ ions, requiring the activity of the Ca2+ ion channel TRPML3. Previous studies have shown elevated cytosolic Ca2+ concentration in ORF3a-expressing cells. Furthermore, the TRPML3 ion channel is proven to be crucial to the regulatory function of ORF3a as knocking down TRPML3 dramatically reduced the enhanced lysosomal exocytosis in ORF3a-expressing cells. Therefore, ORF3a is not self-sufficient in increasing lysosomal Ca2+ and driving the fusion of lysosomes with the PM. The role of TRPML3 is paramount for ORF3a to enhance the process of lysosomal exocytosis (Chen et al. 2021).

However, the mechanism and evolutionary significance of the two aforementioned pathways are yet to be studied. For instance, even though the function of ORF3a inhibiting autophagosome maturation has been elucidated (Chen et al. 2021), there has not yet been a cross-analysis between the ORF3a of different variants and how they may contribute to differentiated transmissibility and virology. How ORF3a mutations may affect its affinity to the VPS39 protein component, and how that may account for differences between the variants is the key point of this research. The mechanism underlying the activation of TRPML3 in the ORF3a-induced enhancement of lysosomal exocytosis is yet to be investigated (Chen et al. 2021). Therefore, this study aims to provide a proposal to answer the two questions from a bioinformatics perspective and an experimental perspective, respectively.

2. Methodology

2.1. Investigation into ORF3a Affinity to VPS39 in Different Variants

In this approach, genomic analysis and protein docking is used to model the binding of ORF3a and VPS39. By comparing the ORF3a coding region of different variants, this study investigates the genetic differences in different SARS-CoV-2 variants. It is further proposed that the mutations in the ORF3a coding region could be investigated as these mutations may lead to different affinities to the VPS39 component. From this, one of the possible factors contributing to the varying transmissibility of different variants may be suggested.

2.2. Investigation into the Pathway Through Which ORF3a Mediates Lysosomal Exocytosis Through TRPML3 Ion channel

To obtain insights into the mechanism by which ORF3a mediates lysosomal exocytosis by acting through TRPML3, this study aims to identify certain signaling factors or proteins that have differentiated expression levels that are correlated with changes in ORF3a and TRPML3 expression. Previous research has revealed that much of the SARS-CoV-2 proteins form complexes with existing human proteins, such as the ORF3a binding to the HOPS complex, leading to protein-protein interactions (Jahanafrooz et al. 2022). Therefore, this study hypothesizes that there is also a certain protein messenger that plays a role in the signaling between ORF3a and TRPML3.

To elucidate this pathway, this study proposes to create an ORF3a-expressing cell. A microarray assay will be conducted to test for proteins that have increased or decreased expression when ORF3a is present to screen for candidate proteins that might be correlated with ORF3a expression and TRPML3 expression. Subsequently, vectors for these protein-coding genes will be constructed and expressed in cells. Finally, it will be possible to identify which candidate proteins may affect TRPML3 expression by measuring the TRPML3 expression levels.

Subsequently, vectors for these protein-coding genes will be constructed and expressed in cells. Finally, it will be possible to identify which candidate proteins may affect TRPML3 expression by measuring the TRPML3 expression levels.

3. Methods and Materials

3.1. Bioinformatics analysis of ORD3a Binding with VPS39

3.1.1. Collection of Viral Genomes of Different Variants

The Refseq, or the standard sequence, for ORF3a protein coded YP_009724391, was used as the reference (wild type) sequence. The genomes of five variants were downloaded, each being the B.1.1.7 (Alpha), the B.1.351 (Beta), the P.1 (Gamma), the B.1.617.2 (Delta), and the B.1.1.529 (Omicron) variant, from NCBI's SARS-CoV-2 nucleotide records database. Ten samples for each variant were chosen. Each of the 10 samples was reported in different countries at different periods. The number of ambiguous characters (N) was set as 0. Tables were generated in Numbers. (See Appendix)

3.1.2. Phylogenetic Analysis

Phylogenetic analysis was conducted using the neighbor-joining method of the MEGAX software. The Bootstrap replicate was set to 1000. Bootstrapping infers the confidence values of phylogenetic trees based on reconstructing trees called "replicates" from minor variations of the input data. In short, this technique serves to verify the reliability of the generated tree. The closer the Bootstrap value is to 1000, the more confident the clade is. Plots were generated and annotated in iTOL (https://itol.embl.de/).

3.1.3. Pairwise Distance Analysis

A pairwise distance analysis was conducted by dividing the sequences into five groups of their respective variants. Then, using the Multiple Sequence Alignment tool MUSCLE, the groups with default settings were aligned so they could be used in the construction of a phylogenetic tree. The results were then inserted into MEGA-X to compute the Pairwise Distance. The Jones-Taylor-Thornton (JTT) method was utilized to analyze nucleotide substitution with Bootstrap replicate set to 1000. A heat-map was generated using MEGA-X (*Figure 3*).

3.1.4. ORF3a Mutational Analysis

Ten sequences of each variant can be compared with reference to the Wuhan/WIV04/2019-12-30/L ReferenceSequence using tools from NGDC (https://ngdc.cncb.ac.cn/ncov/online/tool/variation). The Genome-to-Variant Tool can be used to help identify the mutation sites on the ORF3a coding region (nucleotide sequence 25393–26220). Then the mutation annotation can be conducted using tools from the National Genomic Data Center (China) (https://ngdc.cncb.ac.cn/ncov/online/tool/annotation) by inserting the SNPs within the coding region that corresponds to ORF3a.

3.1.5. ORF3a and VPS39 Protein Docking

The online tool bioinfo3D (http://bioinfo3d.cs.tau.ac.il/wk/index.php/Main_Page) is used to visualize how mutations in ORF3a can lead to differentiated binding with VPS39 in different variants.

3.2. Experiments Proposed to Identify Protein-TMPRL3 Interactions

3.2.1. Cell Culturing

In this study, the HBEC3-KT cells, an immortalized lung epithelial cell line, will be used. It is the in-vivo target cell of SARS-CoV-2. Cells will be obtained from American Type Culture Collection (ATCC: serial number CRL-4051). The growth medium for the cell will be Airway Epithelial Cell Basal Medium (ATCC PCS-300-030) supplemented with the Bronchial Epithelial Cell Growth Kit (ATCC PCS-300-040), both can be obtained from ATCC. All samples should be kept at 37˚C, with 5% CO2.

3.2.2. ORF3a Expression In-Vivo

SARS-CoV-2 ORF3a genes will be amplified by PCR and cloned into pcDNA6B plasmids. The pcDNA6B vector is designed for the overproduction of recombinant proteins in mammalian cell lines. E. coli strain TOP10F is used for the growth of the vector. DNA transfections will be performed with Lipofectamine 2000 for 24h, and ORF3a siRNAs (GenePharma) will be transfected with Lipofectamine RNAiMAX

(13778150, Invitrogen) for 72h. The expression of ORF3a in vivo is verified by Western blot. SARS-CoV-2 ORF3a antibody #34340 can be purchased from Cell Signaling Technology. The dilution ratio should be 1:10000.

3.2.3. DNA microarray Assays

mRNA samples of ORF3a-expressing cells will be collected. A labeling mix containing poly-T primers, reverse transcriptase, and fluorescent-dyed nucleotides will be added to the RNA. PEPperCHIP® Discovery Microarray plates will be used in this investigation.

3.2.4. Patch Clamping

To track how the movement of Ca2+ ions differentiate in ORF3a expressing and normal epithelial cells, a Manual Patch Clamping (MPC) will be performed. Ca2+ currents will be measured in the whole-cell patch clamp configuration (as shown in Figure 1d), which can record currents through multiple channels simultaneously. Voltage-clamp signals will be recorded using an amplifier (Axopatch 200B) connected to a digital interface, (Digidata 1440A) and analyzed using pCLAMP 9 software (Axon Instruments Inc., Burlingame, CA, USA). Pipettes will be pulled from borosilicate glass capillaries (GC150-TF10; Clark Electromedical Inc., Reading, UK) and connected to the head stage of the patch clamp amplifier. The resistance of pipettes in the bath solution will range between 4 and 6 M.

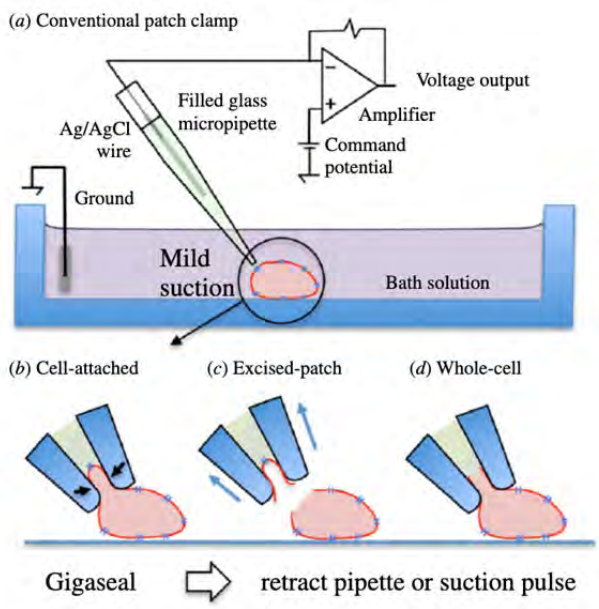


Figure 1. *Cross-sectional schematics describing the patch clamp technique. (a) The conventional approach is routinely performed by utilizing a glass micropipette electrode on a cell adhered to solid support arranged in various recording configurations: (b) cell-attached, (c) excised-patch and (d) whole-cell mode (Yobas 2013).*

Internal and external bath solutions contained (in mM): 145 NaCl, 4 CsCl, 1 CaCl2, 1 MgCl2, 10 glucose, and 10 TES titrated with NaOH to pH 7.4. This is to ensure the external environment of the cell is similar to that in the cytoplasm.

3.2.5. Detection of TRPML3 Expression Levels

mRNA of ORF3a expressing cells will be extracted and reverse transcribed into cDNA. Then theTRPML3 gene will be amplified and recorded using RT-qPCR.

3.2.6. Vector Construction and Transfection

Protein mRNAs that have high fluorescent intensity identified in 3.2.3 will be amplified using PCR. Their respective sequences will also be recorded. Primers will be designed for the mRNA expressing these peptides. After yielding the cDNA sequences of the designated peptides, they will be amplified using PCR and recombined into plasmids. The recombinant plasmid will be transferred to HBEC3-KT through electroporation.

3.2.7. Real-time Quantitative PCR of TRPML3

After being expressed in HBEC3-KT cells, the mRNA will be extracted. For cDNA synthesis, 1µg of total RNA or RNA treated with RNAse free DNase (Invitrogen) will be reverse transcribed. GAPDH will be used as internal control for RT-qPCR since its expression remains constant in the HBEC3-KT cells. TRPML3 cDNA will be amplified through RT-qPCR. Data can be obtained by plotting the Ct values against the log initial concentration.

4. Results and Discussion

4.1. Results

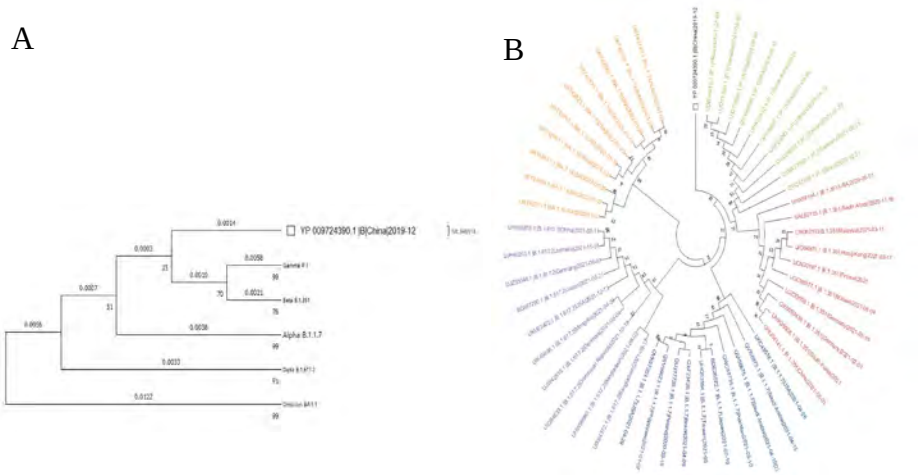


Figure 2. *(A) Neighbor-Joining Phylogenic Tree based on the different variants, showing its evolutionary distance and bootstrap value. (B) Neighbor-Joining Phylogenic Tree based on all samples from all variants.*

Overall, the variants that developed earlier in the pandemic revealed closer evolutionary distance with the original reference sequence NC_045512.2, which as expected illustrates that the evolution has made the sequences more and more different from the original sequence during the past two years of the pandemic. In *Figure 2A*, the Beta (B.1.351) variant has the closest evolutionary distance from the original sequence. It has an evolutionary distance of 0.031. As a variant that follows shortly after, Gamma (P.1) variant is also revealed to have a closer evolutionary distance compared with the Beta (B.1.351) variant. The two closely related variants are also the most closely related to the reference sequence. By contrast, the Delta (B.1.617.2) and Omicron (BA.1.1) variants are shown to have the furthest evolutionary lineage when compared to the reference sequence. However, the Alpha (B.1.1.7) variant, being the earliest VOC, has revealed an evolutionary distance that is further away from the reference sequence compared to the Beta (B.1.351) or Gamma (P.1) variant.

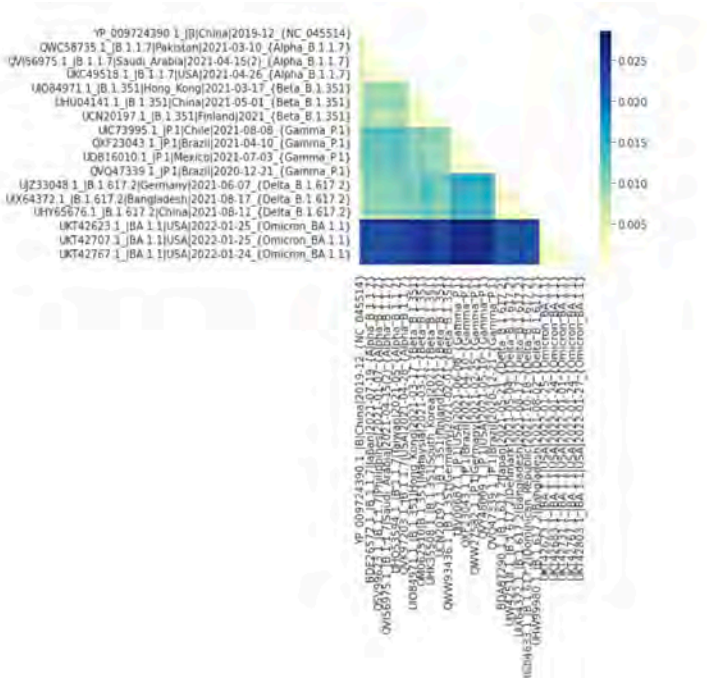


Figure 3. *Heat Map illustrating the pairwise evolutionary distance of all samples from different variants including the reference sequence YP_009724390. The darker color indicates a further evolutionary distance from the reference sequence.*

Similarly, *Figure 3* also reinforces the conclusion that the Omicron variant has the furthest evolutionary distance compared to all other variants. It has a pairwise evolutionary distance value of over 0.0025 when compared to all other variants which have distances of 0.0015 to 0.0020. This indicates that the Omicron variant has experienced the greatest number of mutations, possibly accounting for its increased transmissibility, shorter time to pathogenicity, and decrease in severity.

Furthermore, future mutational analysis could shed light on how certain

mutations may lead to differential transmissibility and pathogenicity. Similarly, the cross-analysis of ORF3a mutation sites in different variants may also reveal how certain mutations could lead to differentiated characteristics between different variants. For instance, certain mutations may be partially accountable for the increased transmissibility of the Omicron (BA.1) variant.

Subsequently, protein docking of ORF3a protein with the VPS39 component of the HOPS complex could verify the previous assumptions about how the transmissibility of different variants may be affected by OFR3a mutations. For example, if ORF3a mutations lead to higher binding affinity to VPS39, then it indicates that the mutation is beneficial for the viral strain as it has a better effect in preventing autophagosomes from binding to lysosomes. This then indirectly increases the efficiency of viral exit as there would be more available lysosomes.

Thus, the impact of this result is two-fold. First, it may reveal the correlation between the pathogenicity or transmissibility of different variants and the mutations that have a high frequency of appearing. Second, it correlates the change in ORF3a affinity to VPS39 caused by these mutations and, therefore, paves the way to help predict the possible outcomes of certain mutations.

4.2. Discussion

The second part of this study investigates the signaling mechanism by which the ORF3a protein interacts with the TRPML3 ion channel to create conditions for lysosomal exocytosis. Expressing only ORF3a in vivo instead of infecting cells with SARS-CoV-2 eliminates possible interferences due to the interactions between other viral proteins and human cell signaling components. Therefore, by only transfecting cells with ORF3a, it can be verified that any changes in protein expression could be attributed to ORF3a.

TRPML3 is among three of the evolutionarily conserved, non-selective cation channels expressed in endolysosomal vesicles along with TRPML1, 2. While TRPML1 is expressed ubiquitously in all tissues, TRPML3 and TRPML2 are found in more specialized cell types—signifying the functional specificities of these isoforms (García-Añoveros and Wiwatpanit 2014). TRPML3 is mostly found in, endocrine, kidney, and lung organs but also in immune cells, while TRPML2 is predominantly expressed in the thymus, spleen, and immune cells (Spix et al. 2020). Its function as a Ca2+ permeable channel helps in regulating endocytosis, phagocytosis, and most importantly lysosomal exocytosis of materials. In the process of lysosomal exocytosis, the regulation of Ca2+ concentration is important. Increased Ca2+ concentration typically increases the secretion of materials by 10-15% (Rosato et al. 2021). In viral-related exocytosis, it has been demonstrated that when the TRPML3 is absent or inhibited, the exocytosis of several viruses such as the MERS-CoV, SARS-CoV, Ebola virus, influenza A virus, and yellow fever virus, are blocked or slowed down (Grimm and Tang 2020). Therefore, TRPMLs is suggested as a potential target for the treatment of viral infectious diseases. However, the mechanism by which the cell signals the activation of TRPML3 remains unclear. Therefore, this study contributes to proposing an experiment to verify the mechanism in which SARS-CoV-2 utilizes human proteins to increase TRPML3 expression, ultimately contributing to increased viral release. In the end, the results of this experiment may serve as the basis for therapeutic treatment of not only SARS-CoV-2 but also many of the aforementioned viruses. By identifying the protein pathway through which TRPML3 is signaled, future investigations may work towards inhibiting the pathway to serve as a means of treatment.

To reinforce the results of the proposed experiment, the microarray test for ORF3a-expressing cells will contribute towards identifying which proteins have an increased expression when compared to the control. On the other hand, the detection of TRPML3 expression levels in ORF3a-expressing cells is to further elucidate the role of ORF3a in increasing TRPML3 expression. Then, peptides that were revealed to have increased expression levels in the microarray test would be isolated and amplified to create a vector. They can then be transfected into epithelial cells, then again, the expression levels of TRPML3 can be tested to further elucidate if these peptides do play a role in the activation of TRPML3 channels.

References

Antonin W. 2000. A SNARE complex mediating fusion of late endosomes defines conserved properties of SNARE structure and function. *The EMBO Journal*. 19(23):6453–6464. doi:10.1093/emboj/19.23.6453.

Bianchi M, Borsetti A, Ciccozzi M, Pascarella S. 2021. SARS-Cov-2 ORF3a: Mutability and function. *International Journal of Biological Macromolecules*. 170 (2021):820–826. doi: 10.1016/j.ijbiomac.2020.12.142.

Chen D, Zheng Q, Sun L, Ji M, Li Y, Deng H, Zhang H. 2021. ORF3a of SARS-CoV-2 promotes lysosomal exocytosis-mediated viral egress. *Developmental Cell*. 56(23). doi:10.1016/j.devcel.2021.10.006.

Ciotti M, Angeletti S, Minieri M, Giovannetti M, Benvenuto D, Pascarella S, Sagnelli C, Bianchi M, Bernardini S, Ciccozzi M. 2020. COVID-19 Outbreak: An Overview. *Chemotherapy*. 64(5-6):1–9. doi:10.1159/000507423.

García-Añoveros J, Wiwatpanit T. 2014. TRPML2 and mucolipin evolution. *Handbook of Experimental Pharmacology*. 222:647–658. doi:10.1007/978-3-642-54215-2_25. [accessed 2022 Aug 30]. https://pubmed.ncbi.nlm.nih.gov/24756724/.

Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, O'Meara MJ, Rezelj VV, Guo JZ, Swaney DL, et al. 2020. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*. 583(459–468). doi:10.1038/s41586-020-2286-9.

Grimm C, Tang R. 2020. Could an endo-lysosomal ion channel be the Achilles heel of SARS-CoV2? *Cell Calcium*. 88(102212):102212. doi:10.1016/j.ceca.2020.102212.

Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu N-H, Nitsche A, et al. 2020. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*. 181(2):271–280. doi:10.1016/j.cell.2020.02.052.

Issa E, Merhi G, Panossian B, Salloum T, Tokajian S. 2020. SARS-CoV-2 and ORF3a: Nonsynonymous Mutations, Functional Domains, and Viral Pathogenesis. Gilbert JA, editor. *mSystems*. 5(3). doi:10.1128/msystems.00266-20. [accessed 2020 Jul 3]. https://msystems.asm.org/content/msys/5/3/e00266-20.full.pdf.

Jahanafrooz Z, Chen Z, Bao J, Li H, Lipworth L, Guo X. 2022. An overview of human proteins and genes involved in SARS-CoV-2 infection. *Gene*. 808:145963. doi:10.1016/j.gene.2021.145963.

Jiang P, Nishimura T, Sakamaki Y, Itakura E, Hatta T, Natsume T, Mizushima N. 2014. The HOPS complex mediates autophagosome–lysosome fusion through interaction with syntaxin 17. Yoshimori T, editor. *Molecular Biology of the Cell*. 25(8):1327–1337. doi:10.1091/mbc.e13-08-0447.

Jungreis I, Sealfon R, Kellis M. 2021. SARS-CoV-2 gene content and COVID-19 mutation impact by comparing 44 Sarbecovirus genomes. *Nature Communications*. 12(1):2642. doi:10.1038/s41467-021-22905-7. https://pubmed.ncbi.nlm.nih.gov/33976134/.

Kadam SB, Sukhramani GS, Bishnoi P, Pable AA, Barvkar VT. 2021. SARS-CoV-2, the pandemic coronavirus: Molecular and structural insights. *Journal of Basic Microbiology*. 61(3):180–202. doi:10.1002/jobm.202000537.

Kern DM, Sorum B, Mali SS, Hoel CM, Sridharan S, Remis JP, Toso DB, Kotecha A, Bautista DM, Brohawn SG. 2021. Cryo-EM structure of SARS-CoV-2 ORF3a in lipid nanodiscs. *Nature Structural & Molecular Biology*. 28(7):573–582. doi:10.1038/s41594-021-00619-0. [accessed 2021 Aug 29]. https://www.nature.com/articles/s41594-021-00619-0.

Kirtipal N, Bharadwaj S, Kang SG. 2020. From SARS to SARS-CoV-2, insights on structure, pathogenicity and immunity aspects of pandemic human coronaviruses. *Infection, Genetics and Evolution*. 85:104502. doi:10.1016/j.meegid.2020.104502.

Lamb CA, Yoshimori T, Tooze SA. 2013. The autophagosome: origins unknown, biogenesis complex. *Nature Reviews Molecular Cell Biology*. 14(12):759–774. doi:10.1038/nrm3696.

Miao G, Zhao H, Li Y, Ji M, Chen Y, Shi Y, Bi Y, Wang P, Zhang H. 2021. ORF3a of the COVID-19 virus SARS-CoV-2 blocks HOPS complex-mediated assembly of the SNARE complex required for autolysosome formation. *Developmental Cell*. 56(4):427-442.e5. doi:10.1016/j.devcel.2020.12.010.

Pu J, Guardia CM, Keren-Kaplan T, Bonifacino JS. 2016. Mechanisms and functions of lysosome positioning. *Journal of Cell Science*. 129(23):4329–4339. doi:10.1242/jcs.196287.

Redondo N, Zaldívar-López S, Garrido JJ, Montoya M. 2021. SARS-CoV-2 Accessory Proteins in Viral Pathogenesis: Knowns and Unknowns. *Frontiers in Immunology*. 12. doi:10.3389/fimmu.2021.708264.

Rosato AS, Tang R, Grimm C. 2021. Two-pore and TRPML cation channels: Regulators of phagocytosis, autophagy and lysosomal exocytosis. Pharmacology & Therapeutics. 220:107713. doi:10.1016/j.pharmthera.2020.107713.

Saftig P, Klumperman J. 2009 Sep 1. Lysosome Biogenesis and Lysosomal Membrane Proteins: Trafficking Meets Function. *Nature reviews Molecular cell biology*. https://pubmed.ncbi.nlm.nih.gov/19672277/.

Spix B, Chao Y-K, Abrahamian C, Chen C-C, Grimm C. 2020. TRPML Cation Channels in Inflammation and Immunity. *Frontiers in Immunology*. 11. doi:10.3389/fimmu.2020.00225.

World Health Organization. 2022. WHO COVID-19 dashboard. World Health Organization. https://covid19.who.int/.

Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, Hu Y, Tao Z-W, Tian J-H, Pei Y-Y, et al. 2020. A new coronavirus associated with human respiratory disease in China. *Nature*. 579(7798). doi:10.1038/s41586-020-2008-3.

Wu P-H, Onodera Y, Giaccia AJ, Le Q-T, Shimizu S, Shirato H, Nam J-M. 2020. Lysosomal trafficking mediated by Arl8b and BORC promotes invasion of cancer cells that survive radiation. *Communications Biology*. 3(1):1–15. doi:10.1038/s42003-020-01339-9. [accessed 2022 Aug 3]. https://www.nature.com/articles/s42003-020-01339-9.

Yobas L. 2013. Microsystems for cell-based electrophysiology. *Journal of Micromechanics and Microengineering*. 23(8):083002. doi:10.1088/0960-1317/23/8/083002.

Zhang J, Ejikemeuwa A, Gerzanich V, Nasr M, Tang Q, Simard JM, Zhao RY. 2022. Understanding the Role of SARS-CoV-2 ORF3a in Viral Pathogenesis and COVID-19. *Frontiers in Microbiology*. 13. doi:10.3389/fmicb.2022.854567.

Zhao YG, Zhang H. 2019. Autophagosome maturation: An epic journey from the ER to lysosomes. *The Journal of Cell Biology*. 218(3):757–770. doi:10.1083/jcb.201810099. [accessed 2022 Feb 14]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6400552/.

Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, et al. 2020. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine*. 382(8). doi:10.1056/nejmoa2001017.

Appendix

Table 1. *Virus variant, NCBI accession, collection date, pangolin lineage, and country of collection. The 50 sequences are classified into groups of 10 based on their pangolin lineage. These sequences are used in both phylogenic studies and pairwise distance analysis.*

Sample	NCBI Accession	Collection Date	Pangolin	Country
Alpha	MZ477832	2021-04-09	B.1.1.7	Brazil
	BS001491	2021-07-19	B.1.1.7	Japan
	MZ328042	2021-03-10	B.1.1.7	Pakistan
	MW735424	2021-01-07	B.1.1.7	Philippines
	MZ047082	2020-03-15	B.1.1.7	Poland
	MZ208928	2021-04-15	B.1.1.7	Saudi Arabia
	OV104898	2021-04-19	B.1.1.7	Slovakia
	OM021311	2021-05	B.1.1.7	Taiwan
	OM486645	2021-04-26	B.1.1.7	USA
	MZ226112	2021-04-28	B.1.1.7	USA
Beta	OM463433	2021-03-16	B.1.351	Germany
	OM212470	2021-03-17	B.1.351	Hong Kong
	OM095211	2020-05-01	B.1.351	USA
	OM062510	2021-03-11	B.1.351	Malaysia
	OM062573	2021-05-01	B.1.351	China

Sample	NCBI Accession	Collection Date	Pangolin	Country
	OL966993	2021	B.1.351	South Korea
	OL779034	2021-08-04	B.1.351	Malawi
	OK448476	2021	B.1.351	Finland
	OK091660	2020-11-16	B.1.351	South Africa
	MZ433432	2021-02-01	B.1.351	Germany
Gamma	OM146081	2021-08-08	P.1	Chile
	OM433396	2021-06-08	P.1	USA
	OM367886	2021-04-20	P.1	Canada
	MZ477800	2021-04-10	P.1	Brazil
	OL966995	2021	P.1	South Korea
	MZ427312	2021-02-25	P.1	Germany
	OK550275	2021-07-03	P.1	Mexico
	MZ310264	2021-05-10	P.1	USA
	MZ277388	2021-01-28	P.1	Taiwan
	MZ264787	2020-12-21	P.1	Brazil
Delta	OM653624	2021-12-13	B.1.617.2	USA
	LC646473	2021-05-21	B.1.617.2	Japan
	OM463389	2021-06-07	B.1.617.2	Germany
	OM443077	2021-05-04	B.1.617.2	Denmark
	OM320390	2021-11-01	B.1.617.2	Jamaica
	OM277522	2021-08-17	B.1.617.2	Bangladesh
	OM190671	2021-04-24	B.1.617.2	Mongolia
	OM180371	2021-10-18	B.1.617.2	Dominican Republic
	OM108132	2021-08-11	B.1.617.2	China
	OM090150	2021-08-02	B.1.617.2	Bangladesh

Sample	NCBI Accession	Collection Date	Pangolin	Country
Omicron	OM646995	2022-01-25	BA.1.1	USA
	OM646996	2022-01-25	BA.1.1	USA
	OM646997	2022-01-01	BA.1.1	USA
	OM647001	2022-01-24	BA.1.1	USA
	OM647003	2022-01-25	BA.1.1	USA
	OM647005	2022-01-01	BA.1.1	USA
	OM647006	2022-01-02	BA.1.1	USA
	OM647008	2022-01-24	BA.1.1	USA
	OM647010	2022-01-18	BA.1.1	USA
	OM647011	2022-01-27	BA.1.1	USA
NCBI Reference Sequence	NC_045512.2	2019-12-30	B	China

Localising the Source of Calcium for Slow Adaptation in Cochlear Hair Cells

Shuhan Cao

Author Background: *Shuhan Cao grew up in Hong Kong and currently attends German Swiss International School in Hong Kong, China. Her Pioneer research concentration was in the field of biology/neuroscience and titled "Modern Topics in Sensory Neurobiology."*

Abstract[1]

Hair cells located in the cochlea detect stimuli through mechanoelectrical transduction (MET) channels, which are pulled open during bundle deflection by tip links which connect a shorter stereocilium to a taller adjacent neighbour. Adaptation is a decline in sensory response to an unchanging stimulus, and is a key feature exhibited by the mechano-electric transduction process that allows the sensory system to remain sensitive to changes in the environment. Although evidence suggests that calcium plays an important role in modifying slow adaptation, the exact source of calcium remains unknown. MET and voltage-gated calcium channel blockers were used in this study along with high-speed calcium imaging to determine the entry point and source of calcium. Adaptation was found to decrease moderately when blocking MET and voltage-gated calcium channels separately and decrease significantly when blocking both. High-speed calcium imaging showed back diffusion of calcium from the second row of stereocilia to the tallest first row. We hypothesised that the adaptation rate would change with the length of stereocilia, as diffusion time changes with distance. Adaptation rate was found to increase when using the Myo15 mutant, which results in shorter stereociliary length, and decrease when using the Whirler mutant, which results in longer stereociliary length. The time taken for adaptation to occur in these mutants along with the control was fitted to a power function consistent with the Einstein law of diffusion ($r=\sqrt{Dt}$). These observations suggest that calcium required for slow adaptation enters through MET channels and

[1] **Editorial Note:** In the Pioneer research concentration "Modern Topics in Sensory Neurobiology," scholars are tasked with proposing a study to address a yet unsolved question in the field of neurobiology. This paper represents a *hypothetical* research study/experiment, and the results described herein are *hypothesized results* and do not represent the actual results of an executed experimental study.

also voltage-gated calcium channels at the base of stereocilia, and imply that they reach their site of action through diffusion. By increasing the understanding of adaptation, further advances in improving hearing loss can be made.

1. Introduction

Hair cells in the auditory system convert mechanical information from sound into electrical information through stereocilia (Caprara et al., 2020). Stereocilia are actin-filled structures similar to microvilli and are the main components of a sensory hair bundle (Maoiléidigh & Ricci, 2019). They are arranged in rows of increasing height and are connected by a filamentous tip link (Caprara et al., 2020). The tip-link connects the top of one stereocilium to its taller neighbour, and gates mechano-electrical transducer (MET) channels located at the bottom of tip-links (Beurg et al., 2009). Deflection towards the tallest row of stereocilia is excitatory and opens MET channels by increasing the tension on the tip-link (Beurg et al., 2009). Open MET channels allow cations to pass through and depolarise the hair cell, resulting in the release of neurotransmitters onto the afferent auditory neurons (as reviewed in Fettiplace, 2017). Adaptation is a key mechanism of hair cell mechanotransduction and is a decline in sensory response to an unchanging stimulus (as reviewed in McPherson, 2018). It allows the sensory system to remain sensitive to differences in the environment. At the physiological level, adaptation is also essential for the sensory system to filter stimuli and suppress background noise, allowing for focus and decreasing distractions. It prevents cortical overstimulation by closing the MET channels, therefore preventing the constant release of neurotransmitters, and has been proposed to help achieve efficient coding of incoming auditory information (as reviewed in Pérez-González & Malmierca, 2014). This study focuses on an element of adaptation known as slow adaptation, and specifically on the source of calcium which drives it.

There are two types of adaptation: fast adaptation and slow adaptation, with fast adaptation having a time constant of less than 10 ms, and slow adaptation having a time constant of 10 ms or more (Caprara et al., 2020). Slow adaptation has been studied for many years, and has been shown to need the entry of calcium ions into the stereocilia and myosin motor activity (Corns et al., 2014), thus leading to the creation of the widely accepted “motor model.” In the motor model, it is hypothesised that myosin motors are attached to the upper end of the tip-link and climb up the actin filaments toward the top of the stereocilia to generate resting tip-link tension. During hair bundle deflection, calcium enters and Ca^{2+}-bound calmodulin is hypothesised to cause the slipping of the myosin motor down the stereocilium side to reduce tip-link tension, thus closing the MET channels and resulting in adaptation (Caprara et al., 2020). However, despite extensive research being done in this area, the exact source of calcium for slow adaptation remains unknown (Caprara et al., 2020). Previously, it had been thought that calcium entered through MET channels located at the upper insertion site of the tip-link to interact with the myosin, but evidence now calls into question the validity of this, as there are no MET channels at the upper end

of tip-links (Beurg et al., 2009).

Not only do the MET channel positions call into question the motor model, but the specific type of myosin has also been debated. Myosins are a superfamily of motor proteins (*What Is Myosin?*, n.d.) responsible, most notably, for muscle contraction ("Myosin," 2022). It has long been hypothesised that myosin plays a role in adaptation (Gillespie, 2004), usually with myosin 1c being the likely motor located at the top of the tip-link insertion site. Myosin 1c is thought to be responsible for interacting with calcium to result in slow adaptation (as reviewed in Gillespie & Cyr, 2004). However, recent evidence calls into question the role of myosin 1c in adaptation, as experiments have demonstrated the presence of slow adaptation even when myocin 1c is inhibited (Caprara et al., 2020). Consequently, it is now thought that another type of myosin, myosin VIIa, multiple isoforms of which are found in the mice cochlea, may be responsible for tensioning the MET channel complex (Li et al., 2020).

Despite the uncertainties regarding the types of myosin responsible for the motor, the idea that calcium interacts with myosin to cause slow adaptation has been supported. In this study, the source of calcium for slow adaptation in cochlear hair cells was investigated by blocking MET and voltage-gated calcium channels respectively. We hypothesised that if calcium is coming through open MET and voltage-gated calcium channels, then blocking these channels should reduce adaptation. We also used high-speed calcium imaging to track the movement of calcium through the stereocilia by diffusion.

As diffusion depends greatly on distance, we hypothesised that if calcium reaches the myosin motor by diffusion, the rate of adaptation should be faster when shorter stereocilia are used and vice versa. Thus, the Myo15 mutant stereocilia, which are significantly shorter than normal stereocilia, and the Whirler mutant stereocilia, which are significantly longer than normal stereocilia, were employed to investigate the effect of diffusion on slow adaptation. The hair bundle was deflected using a fluid jet to ensure uniform deflection. We found that slow adaptation was considerably decreased when blocking MET or voltage-gated calcium channels, and significantly decreased when blocking both. The calcium markers along with high-speed calcium imaging indicated that there was a back diffusion of calcium from the middle stereocilia row to the tallest row. This is further demonstrated by a faster rate of adaptation when using the Myo15 mutant stereocilia, and a slower rate of adaptation when using the Whirler mutant. These data then support the conclusion that calcium for slow adaptation enters through open MET and voltage-gated calcium channels, and that it travels to their site of action through diffusion. This undoubtedly furthers the current understanding of the auditory system.

2. Materials and Methods

2.1. Electrophysiology

Inner hair cells (n=30) and outer hair cells (n=92) from the mouse cochlea were studied in dissected Organs of Corti. Animals of both sexes were killed by decapitation. Whole-cell patch-clamp recordings of current were performed at room temperature (Corns et al., 2014). Whole-cell patch-clamp recordings were made using an Axon 200B amplifier or a MultiClamp 700B amplifier (Caprara

et al., 2020). Hair bundles were depolarised with a voltage clamp. Patch pipettes were filled with an intracellular solution containing 125 mM CsCl, 3.5 mM MgCl2, 5 mM adenosine triphosphate (ATP), 5 mM creatine phosphate, 10 mM Hepes,, 3 mM ascorbic acid, pH 7.2, 280 to 290 mOsm (Caprara et al., 2020). The figures show the standard error of mean (SEM) in error bars.

2.2. Hair Bundle Stimulation

The following technique is modified from Corns et al., 2014. A fluid jet from a pipette driven by a piezoelectric disk was used to elicit MET currents. The pipette was positioned around 8 micrometres away from the hair bundle. The width of the hair bundle was: 8.7 ± 0.4 µm (n = 10) for IHCs and 6.7 ± 0.1 µm (n = 10) for OHCs (Corns et al., 2014). Inner hair cells are believed to be stimulated by fluid movements in the endolymph, making the fluid jet the most suitable stimulation method as it most accurately replicates the natural stimulation of inner hair cells. The fluid jet also has considerable advantages over the commonly used stiff probe, namely that with the stiff probe it is difficult to ensure uniform deflection of the hair bundle unless the shape of the probe exactly matches the shape of the stereocilia (Corns et al., 2014).

The following technique is modified from Beurg et al., 2009. Mechanical stimuli were applied to displace the hair bundle using the fluid jet, and a patch pipette was inserted into the cell to measure the change in current over time of the cell and investigate the course of slow adaptation (Beurg et al., 2009).

2.3. Hair Bundle Stimulation

The following technique is modified from Beurg et al., 2009. To block the MET channels, a puff pipette was filled with streptomycin (1 mM, N = 4) and the contents were puffed onto the stereocilia at a distance of 25 micrometres. In high-speed calcium imaging, the area outside the hair cell was filled with calcium markers from the Fluo4 family, before being imaged using a swept field confocal lens paired with a high-speed camera (Beurg et al., 2009), which took images every 2ms.

The following technique is modified from Lee et al., 1999. To block the voltage-gated calcium channels, a 100 mM $NiCl_2$ stock solution (stored at room temperature) was used for dilutions in deionised water. The solution was diluted to 300µM. In the experiment, the puff pipette was filled with $NiCl_2$ solution (300 µm) and puffed onto the stereocilia at a distance of 20 micrometres (Lee et al., 1999).

2.4. Auditory Brainstem Response

The following technique is modified from Akil et al., 2016. Tmc1 pD569N homozygote mice and the control group were anaesthetised with an intraperitoneal injection of a mixture of ketamine hydrochloride (Ketaset 100 mg/kg) and xylazine hydrochloride (Xyla-Ject 10 mg/kg). The anesthetization was done using a 1 ml insulin syringe with the precision glide needle. The mouse was placed on a pre-heated mat of approximately 37˚C within a sound-proof chamber and the speaker is placed 10 cm from its left ear. Electrodes were

inserted subdermally at the forehead, pinna of the left ear and below the contralateral (right) ear. The sound-proof chamber was then closed and sounds were presented.

All procedures and animal handling described in this protocol were done according to approved national ethical guidelines and complied with all protocol requirements of the Institutional Animal Care and Use Committee (Akil et al., 2016).

2.5. Data Analysis

The following procedure is modified from Caprara et al., 2020. Activation curves were generated using the displacement values when the peak current occurred for 50-ms step traces. Normalised currents (I/Imax) were generated by subtracting leak current, which is the smallest remaining current during the negative steps and normalising to the peak current. Activation curves were fitted with a double Boltzmann equation, and for mechanical stimulus steps, adaptation time constant fits were obtained at ~50% peak current using a double exponential equation. Data was analysed using Excel (Microsoft) (Caprara et al., 2020). Graphs were plotted and analysed using Graphpad Prism and Desmos, with error bars being the Standard Error of Mean (SEM).

3. Results

When delivering a mechanical stimulus with a fluid jet, the inner hair cell's current is shown to peak and then decay, an indication of the slow adaptation process (Caprara et al., 2020). Adaptation is manifested in electrophysiological recordings through the peak and decline of the current measured. This decrease in current measured demonstrates the closing of the MET channels in response to an unchanging stimulus, the hallmark of adaptation. Despite extensive experiments, the source of calcium remains unknown. In our study, we hypothesised that the calcium required for adaptation enters through both open MET and voltage-gated calcium channels and that they reach their site of action by diffusion. We also hypothesised that the calcium ions entering through the voltage-gated calcium channels and open MET channels interact with the myosin at the top of the tip-link insertion site. This causes the myosin to slip down, orchestrating fast and slow adaptation, and closing MET channels.

3.1. Adaptation Decreases with Blocked MET Channels

To investigate whether the calcium causing slow adaptation is coming through MET channels which are open due to the mechanical stimulus, we used MET channel blockers, including the aminoglycoside antibiotic, streptomycin (Beurg et al., 2009). We made recordings of the transducer current using a patch pipette. Upon using the blockers, a decrease in both adaptation and the rate of adaptation was observed compared with the control (Fig 1a). This suggests that the calcium required for adaptation was partially entering through open MET channels. However, as adaptation was still occurring, this finding also suggests that there are other sources of calcium.

To ensure that the blockers were able to block calcium from coming in,

the cell was injected with calcium marker Fluo4 using a patch pipette, and high-speed calcium imaging was carried out. To prevent a motion artefact due to the mechanical stimulus, an experiment protocol much like the one in Beurg et al., 2009 was adopted (Fig 1b), where the bundle deflection was combined with a depolarising voltage step, so there would be no calcium entering until the cell was repolarised to its normal holding potential. In the control experiment, a burst of fluorescence was observed when the hair bundle was repolarised. However, when using channel blockers, there was significantly less fluorescence (Fig 1c), showing that the blockers successfully blocked the majority of MET channels and prevented calcium from entering the stereocilia through it. Nonetheless, the blockers may have also had other effects on the MET channels that affected the results.

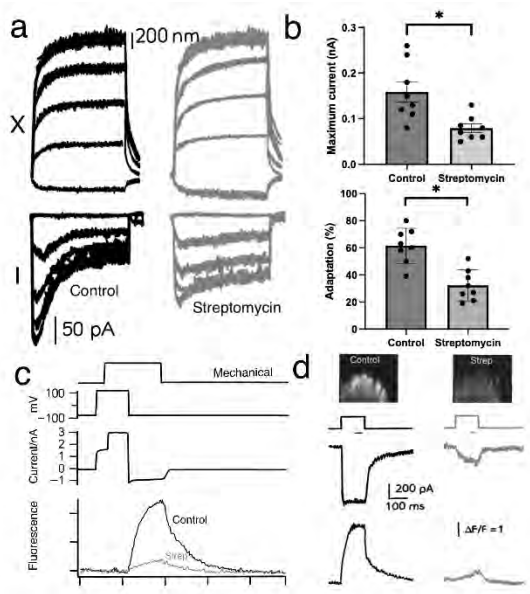


Figure 1: Effect of MET channel blockers on slow adaptation. *(a) Hair bundle displacements (X) and currents (I) were recorded. Slow adaptation decreases when MET channels are blocked with streptomycin, seen through the lower rate of decay of the current measured (grey), but is still occurring. (b) Comparisons between maximum current and % adapted of the control and streptomycin experiment are shown. Streptomycin led to a lower maximum current (top graph) as well as a lower % adapted (lower graph) compared with the control experiment. Error bars are SEM. (c) The protocol being used to prevent a motion artefact when utilising high-speed calcium imaging (upper graphs). The depolarisation from -80 to 100 mV is followed by a bundle stimulus using a fluid jet (mechanical stimulus). The lower graph shows the difference in fluorescence intensity between the control and streptomycin experiments. (d) High-speed calcium imaging shows that streptomycin effectively blocks calcium entry, seen through the differences in fluorescence of the stereocilia between the control and streptomycin experiments (top image). Lines underneath the mechanical stimulus graph indicate the time when the image was taken. The differences in current are seen in the middle two traces, with the streptomycin experiment having a much lower current compared with the control. There was much less fluorescence in the streptomycin experiment, as shown in the bottom traces. However, streptomycin still allows some calcium through as there is still some slight fluorescence. Diagrams modified from Beurg et al., 2009, Caprara et al., 2020.*

3.2. Adaptation Significantly Decreases when Depolarising Cells to a Positive Potential

To ensure that the MET channel blockers were not interacting with the stereocilia in ways that would affect the result of the previous experiment, the cell was depolarised (Fig 2a) to positive potentials close to E_{Ca} to minimise the driving force and consequently decrease the amount of calcium entering through the

MET channels, even when open. This technique has the same effect as blocking MET channels, so we hypothesised the results to be similar to those in Figure 1. When doing this, we found that adaptation was significantly decreased, similar to the results in Figure 1 (Fig 2b). Similarly, when hyperpolarising the cell so the driving force of calcium is increased, increased rates of adaptation are expected, and that is what is found. Comparisons between the percentage adapted and maximum current of the control experiment and when the experiment where the cell was depolarised show less adaptation with the depolarised cell (Fig 2c). This evidence once again suggests that the calcium entering through the MET channels plays an important role in adaptation.

However, like in Figure 1, although adaptation was decreased, it was not completely abolished, and we hypothesised that this may be because the E_{Ca} in the upper and lower part of the cell are different due to different concentrations of calcium in the endolymph and the extracellular fluid. Consequently, depolarising the cell to a certain voltage would only limit calcium entry from one source, while calcium can still enter from another due to the differences in E_{Ca} between the endolymph and extracellular fluid.

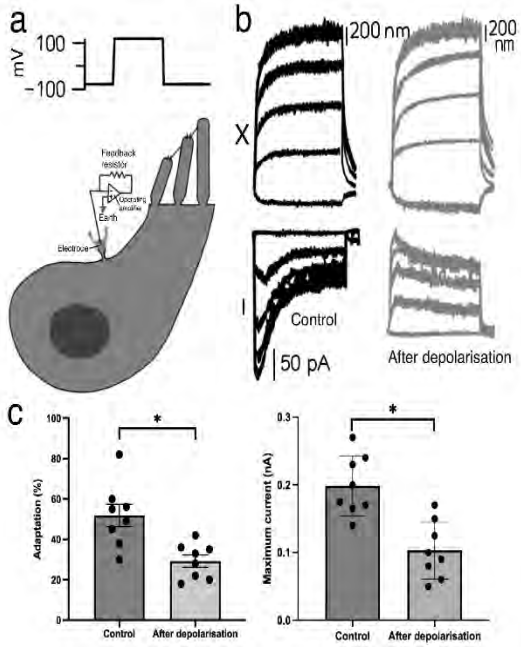


Figure 2: Effects of depolarising the cell on slow adaptation. *(a) The top diagram shows the graph of the depolarisation of the cell to +120mV, close to the calcium equilibrium potential, meaning that no calcium from the endolymph can enter through the MET channels. The depolarisation was done using a patch pipette (bottom diagram). (b) Hair bundle displacements (X) and currents (I) were recorded. The recorded current of the hair bundle when depolarised (right) compared with the control experiment (left) is shown. The current is positive (right, bottom) as potassium is exiting the cell. There is less adaptation exhibited as the rate of decay is lower. (c) The comparison between maximum % adapted (left) and maximum current (right) of the control experiment and experiment after depolarisation. Depolarisation of the cell decreased both the % adapted as well as the maximum current measured. Diagrams modified from Beurg et al., 2009 and Caprara et al., 2020.*

3.3. Adaptation Significantly Reduced when Blocking both MET and Voltage-gated Calcium Channels

Voltage-gated calcium channels and MET channels were blocked to determine whether the results from Figure 2 were because depolarising the cell allows calcium to come in through voltage-gated calcium channels due to differences in the E_{Ca} of the endolymph and extracellular fluid. Voltage-gated calcium channels were first blocked alone using nickel puffed onto the hair bundle with a puff pipette positioned 25 micrometres away from the bundle (Fig 3a), then along with

MET channels using streptomycin. When only blocking voltage-gated calcium channels, adaptation was reduced moderately (Fig 3b). This suggests that while calcium coming from the voltage-gated calcium channels was necessary, they were not the only source of calcium needed for slow adaptation, especially as there was still calcium influx through open MET channels. However, when blocking both MET and voltage-gated calcium channels, slow adaptation was much more significantly reduced compared to when blocking voltage-gated calcium channels alone (Fig 3c). This indicates that these are the main sources of calcium. However, adaptation was still not abolished. It was hypothesised that it was because streptomycin cannot completely block the MET channels and because calcium is always present in the cell in small quantities.

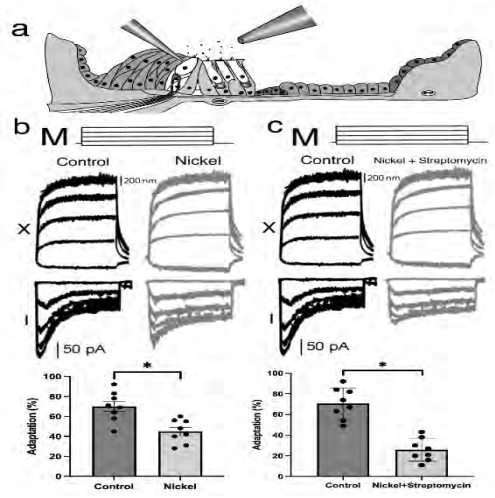


Figure 3: Effects of blocking voltage-gated calcium channels and/or MET channels on slow adaptation *(a) Illustration of the experimental paradigm. (b) The mechanical stimuli applied on the hair bundle when blocking only voltage-gated calcium channels (upper), and the results that were recorded (middle). The lines in black are the control experiment while the grey lines are when the voltage-gated calcium channels are blocked. The % adapted of the control experiment compared to the % adapted when voltage-gated calcium channels are blocked (lower). (c) The mechanical stimuli applied to the hair bundle when blocking both VGCCs and MET 5*

3.4. Back Diffusion of Calcium Observed

From our results, it was hypothesised that if calcium is entering through MET and voltage-gated calcium channels, it is most likely reaching its site of action at the upper insertion site of the tip-link by diffusion. This is because there are no MET or voltage-gated calcium channels directly next to the predicted site of the myosin motor, especially for the tallest row of stereocilia. To test this hypothesis, calcium markers of the Fluo4 family were used and injected into the extracellular solution, and a patch pipette was inserted in the cell to record measurements of the change in the current of the hair bundle over time (Fig 4a). The calcium marker was present only in the extracellular solution, so when the hair bundle was repolarised after the initial depolarisation and mechanical stimulus (Fig 4b), the calcium entering the cell was visible in our high-speed calcium imaging, and we were able to track the movement of calcium through the stereocilia (Fig 4a). We observed that the calcium entering through the MET channels in the lowest and middle rows back diffused into the tallest row of stereocilia, as there were no MET channels there. Calcium influx through voltage-gated calcium channels at the base stereocilia was also seen.

The time taken to diffuse across a distance increases and decreases

exponentially with an increase or decrease in the distance (Einstein, n.d.). Thus, to reinforce our results, stereocilia of the Myo15 mutant were used, which are significantly shorter than normal stereocilia and therefore has a much shorter diffusion distance. We hypothesised that if diffusion is indeed how the calcium reaches the site of the myosin motor, then by using the Myo15 mutant, the rate of adaptation should be increased. When using the Myo15 mutant, we delivered a mechanical stimulus with a fluid jet identical to the one for normal stereocilia. It was observed that the receptor current peaks and then decays faster than normal stereocilia (Fig 4c). This reinforces the idea that calcium reaches the upper insertion site of the tip-link by diffusion. Similarly, when using the Whirler mutant, which results in longer stereocilia, the rate of adaptation is seen to significantly decrease (Fig 4d). When comparing the percentage adapted, the Whirler mutant showed the lowest percentage, while having the highest time constant (Fig 4e). Therefore, this is consistent with our hypothesis that if calcium travels by diffusion, it should take longer to diffuse when there is a longer distance. When plotting the graph of diffusion against distance, the points fit in a power function as predicted in the Einstein law for diffusion (Fig 4f). However, if the calcium's main method of travel to its site of action is diffusion, there would be a delay in the adaptation of the middle row MET channel, but this was not seen. Possible reasons are discussed in the Discussion section.

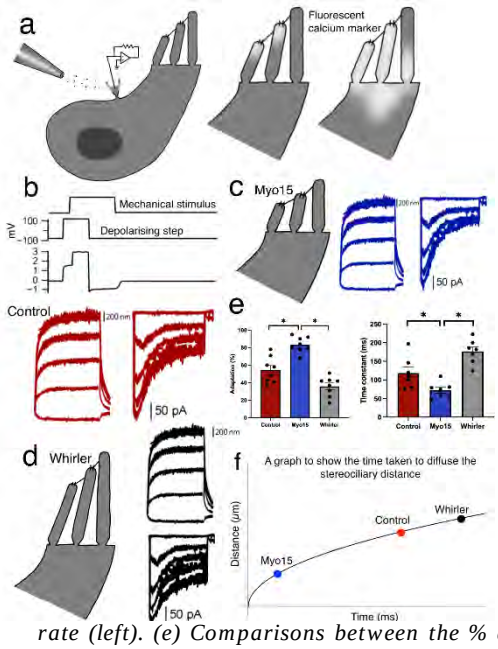


Figure 4: Calcium travels to the myosin motor through diffusion. *(a) A calcium marker of the Fluo4 family is injected into the extracellular solution via a pipette. A patch pipette measures the change in current over time of the hair bundle (left). Calcium (the fluorescence) is shown to diffuse down the first and second rows of stereocilia and into the third row (right). (b) Calcium bound to the marker enters the cell after repolarisation of normal stereocilia (top). Adaptation occurs (bottom). (c) The myo15 mutant, which gives shorter stereocilia, is used (left). The patch pipette measurements (right) show that adaptation happens at a greater rate compared to part b. (d) The Whirler mutant, which gives longer stereocilia is used (left), and adaptation happens at a slower rate (left). (e) Comparisons between the % adapted and time constant of slow adaptation of the normal, Myo15 and Whirler stereocilia. The colours of the dots correspond to the colour of the current traces. (f) The average time to diffuse and cause adaptation is fitted to a power function. Diagrams from/modified from Biorender and Caprara et al., 2020*

3.5. This Mechanism is Conserved in OHCs

Initially, we only investigated the inner hair cells of mice and found from the previous experiments that the calcium required for slow adaptation enters through open MET channels as well as voltage-gated calcium channels, and reaches the upper insertion point of the tip-link through, but is not limited to, diffusion. As the cochlea contains both inner and outer hair cells, we hypothesised that outer hair cells will also have similar results. This is because outer hair cells have a similar structure to inner hair cells. OHCs tune stimuli, amplify sound signals, and also utilise adaptation to filter stimuli (as reviewed in Fettiplace, 2017). When investigating outer hair cells, we found highly similar results to that with inner hair cells (Fig 5), leading us to the conclusion that the main sources of calcium for slow adaptation in both inner and outer hair cells are the MET and voltage-gated calcium channels.

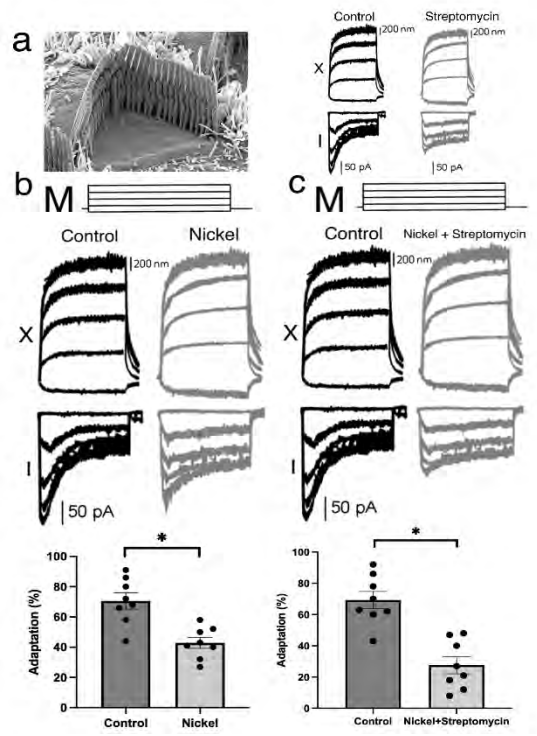


Figure 5: Similar results were found in outer hair cells
(a) Electron micrograph of outer hair cells (left). Current trace when only blocking MET channels with streptomycin (right). (b)The mechanical stimuli delivered to the bundle (upper). Results in outer hair cells after blocking voltage-gated calcium channels (middle). Comparison between % adapted of control experiment and when nickel was used to block the VGCCs (lower). (c) The mechanical stimuli delivered to the bundle (upper). Results/current trace after blocking both voltage-gated calcium channels and MET channels (middle). Comparison between % adapted of control experiment with when both VGCCs and MET channels were blocked. Panel A from Wellcome collection and remaining panels from Caprara et al., 2020

A summary graph of the effects of depolarisation and blockers shows similarities between inner and outer hair cells.

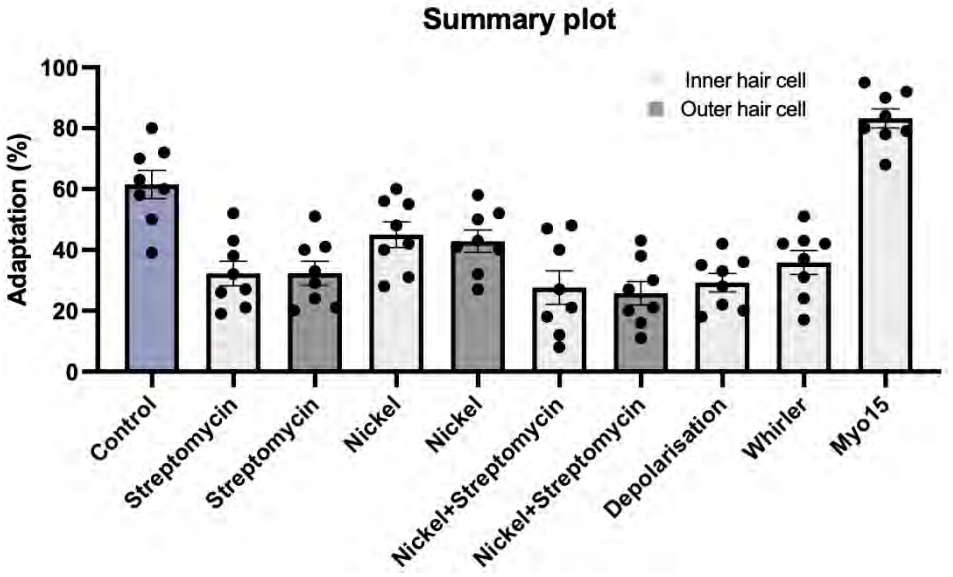


Figure 5a: A summary plot to compare the results between inner and outer hair cells. *The dark grey bars show the results from outer hair cells while the light grey bars show the results from inner hair cells. Results from OHCs and IHCs are placed side by side for easy comparison.*

3.6. Mice with Mutant Stereocilia Show Less Response to Changes in Stimulus

As adaptation has great physiological importance, the next logical step would be to see the effect of inhibiting adaptation to an organism. To investigate the effects of calcium on slow adaptation on the level of the organism, Tmc1 pD569N homozygote mice were used. Tmc1 pD569N homozygote mice have modified MET channels, specifically the protein TMC which makes up the majority of the MET channel, such that the calcium permeability of MET channels is decreased threefold (Beurg et al., 2019, p. 1). Fifteen Tmc1 pD569N homozygote mice were exposed to first a pure tone stimulus (stimulus 1) for three seconds, sufficient time for adaptation to occur, and then a different stimulus (stimulus 2) that should elicit a response in normal mice (Fig 1a). A cat meow was chosen. Behavioural responses and auditory brainstem responses of the Tmc1 pD569N homozygote mice were measured and compared with those of the control group. A positive behavioural response is defined in this experiment as a movement of 4cm or more within one second of switching stimuli.

We found that with the auditory brainstem responses, Tmc1 pD569N homozygote mice showed a significant increase in the second and third curves, which reflects the activity of the VIII cranial nerve and cochlear nucleus (Creel, 2015). As the VIII cranial nerve consists partially of afferent neurons bringing information from the cochlear hair cells (Goutman et al., 2015), overexcitation indicates a lack of adaptation as the continuous influx of cations causes the auditory nerves to fire more frequently (Fig 6b). Not only this, Tmc1 pD569N homozygote mice showed significantly less behavioural response than the control

group (Fig 6c). This shows how calcium entry is not only significant for adaptation but that, as perhaps expected, a decrease in adaptation severely affects an organism's perception of changes in stimuli.

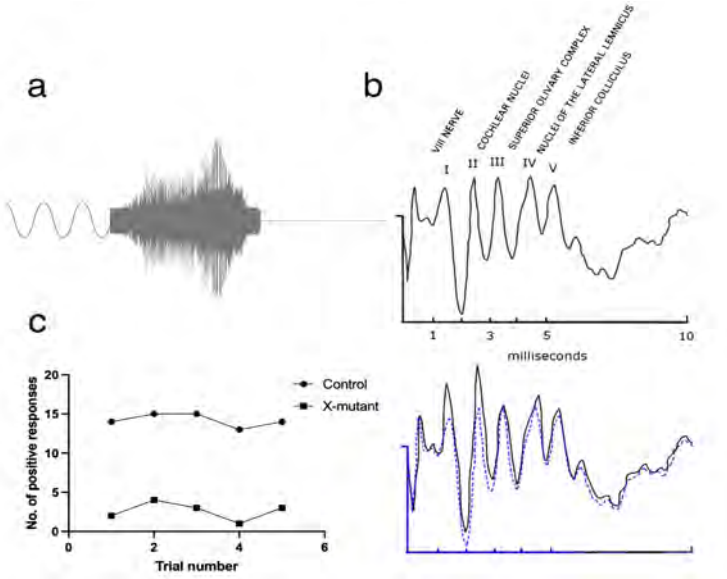


Figure 6: Tmc1 pD569N homozygote mice show less behavioural response and overexcitation of the VIII nerve and cochlear nuclei. *(a) Waveforms of the pure tone stimulus and cat meow stimulus. (b) The auditory brainstem response of the control group (upper; blue, lower) and the ABR of the Tmc1 pD569N homozygote mice (black lower). (c) The comparison between the number of positive behavioural responses of the control group and Tmc1 pD569N homozygote mutant. Diagrams from Fourierstrings, NCBI and samplefocus.*

4. Discussion

In this study, we investigated the sources of calcium required for slow adaptation, focusing mainly on the mouse's inner cochlear hair cells, but also on outer cochlear hair cells. By uniformly deflecting the hair bundle with a piezo-driven fluid jet, we show that adaptation is greatly reduced when blocking MET and voltage-gated calcium channels. These results then indicate that during slow adaptation in both outer and inner hair cells, the main sources of calcium are the open mechanoelectrical transducer channels and voltage-gated calcium channels. Through the experiments, it is also seen through calcium imaging that the calcium reaches the site of the suspected myosin motor through back diffusion. This conclusion is supported by the fact that the rate of slow adaptation is faster when using the stereocilia of the Myo15 mutant, which is significantly shorter than normal, and slower when using the longer Whirler mutant.

The relationship between calcium and adaptation, both slow and fast, has been controversial, with some studies showing that calcium is needed for slow adaptation (Beurg et al., 2008; Corns et al., 2014), and others showing that adaptation happens independent of calcium (Peng et al., 2013). This study reinforces the argument that calcium is a necessary part of adaptation, as we

found that when blocking or reducing calcium entry into the stereocilia, slow adaptation is considerably affected. Adding on to the previous studies regarding the need for calcium for slow adaptation, this paper was able to localise the source of calcium and the method by which calcium reaches the upper insertion point of tip-links, and conclude the sources to be through MET and voltage-gated calcium channels. The method was concluded to be, but not limited to, back diffusion into the stereocilia.

However, there may be still other sources of calcium entry. We found that even when blocking both MET and voltage-gated calcium channels, slow adaptation was still not completely abolished, and this is a possible follow-up study to be done in the future. Perhaps an alternative explanation to why there was still adaptation even though the MET and voltage-gated calcium channels were blocked was that when using the MET channel blocker streptomycin, there were a few channels that were not completely blocked on each occasion (Figure 1c), meaning that this was a possible limitation of our experiment and may have slightly affected our results. Additionally, calcium is always present in small concentrations in the cell, and that may have also caused adaptation to some extent. BAPTA buffering could be used to improve this. Because we hypothesised that calcium reached its site of action through diffusion, we expected a slight delay in the slow adaptation of the middle row MET channel, as it takes time for calcium to enter through and diffuse to the upper tip-link insertion site in the tallest row and cause it to slip down. Although diffusion of calcium was observed, there was no observed delay of slow adaptation in the MET channel at the top of the second stereocilia. This lack of delay suggests that there perhaps may be a more direct source of calcium entry in the tallest row of stereocilia, or that the rate at which the images were taken were too slow to catch the delay. However, more experiments and a better understanding of the upper attachment point of the tip-link are needed to justify this hypothesis.

When tracing the movement of calcium through the stereocilia using high-speed calcium imaging, calcium influx was also seen from voltage-gated calcium channels located at the base of stereocilia. We believe that this is to reduce the time taken for calcium to diffuse to the myosin motor at the top of the tallest stereocilia row. Having voltage-gated calcium channels at the base of the middle row of stereocilia reduces the diffusion distance by half, and this enables the rate of adaptation to increase, thus benefiting the organism, as it allows them to respond more quickly to stimulus changes.

From our experiment using Myo15 and Whirler mutants, the length of stereocilia is seen to play an important role in the rate of diffusion. Even when there are no mutations, the lengths of stereocilia vary naturally throughout the distance of the basilar membrane (as reviewed in McPherson, 2018), which is a stiff structural element within the cochlea (“Basilar Membrane,” 2022). Different areas of the basilar membrane vibrate as a reaction to different frequencies of sound, with the apex being sensitive to low frequencies and the basal end sensitive to high frequencies. Stereocilia located at the apex are longer than stereocilia located at the basal end (as reviewed in McPherson, 2018). Given the effect of stereocilium length on the rate of adaptation, we hypothesise that the difference in length of the stereocilia may be due to the differences in the required rate of adaptation. Sounds of different frequencies have different wavelengths, with wavelengths decreasing as frequency increases. As the frequency is encoded by

the firing rate of neurons (*Perception Lecture Notes: Frequency Tuning and Pitch Perception*, n.d.), a high frequency means that auditory neurons must fire more frequently. Therefore, faster rates of adaptation are needed to enable fast firing and fast response to stimulus changes. Consequently, stereocilia at the basal end are shorter to enable faster diffusion and therefore faster adaptation rate. However, this is currently hypothetical, and more experiments will be needed to justify this prediction, making this a possible area of future research.

Adaptation is an extremely important feature of all sensory systems, as it allows the sensory system to remain sensitive to changes in the environment and filter stimuli. It is adaptation that allows for the sensory system to suppress background noise (Khalighinejad et al., 2019), which otherwise will become a constant distraction and impede focus. The two types of adaptation, fast and slow adaptation, are necessary to prevent the over-excitation of the auditory nerve by limiting the amount of calcium entering through the MET and voltage-gated calcium channels. Should the mechanism of adaptation be hindered in any way, the cochlear nerve will become over-excited and the organism's ability to detect stimulus changes is decreased. This is especially demonstrated in Tmc1 pD569N homozygote mice whose cochlear nerves are shown to be overstimulated in the auditory brainstem response. They were also shown to be unable to respond to stimulus changes due to mutations in their MET channel complex which inhibits the entry of calcium. Therefore, should adaptation in humans be compromised, it will greatly affect their quality of life.

Despite the importance of calcium, the source of calcium entry for adaptation has long remained unknown (Caprara et al., 2020). This study can provide evidence that significant sources of calcium are the MET and voltage-gated calcium channels. This study also advances the understanding of slow adaptation and furthers the understanding of the process of transduction in hair cells. Adaptation is useful in many cases and can be used to, for instance, create better cochlear implants (Azadpour & Smith, 2016). A better understanding of adaptation may be used to further improve the lives of those suffering from hearing disorders, benefiting the population greatly.

5. References

Akil, O., Oursler, A. E., Fan, K., & Lustig, L. R. (2016). Mouse Auditory Brainstem Response Testing. *Bio-Protocol*, *6*(6), e1768. https://doi.org/10.21769/BioProtoc.1768

Azadpour, M., & Smith, R. L. (2016). Enhancing speech envelope by integrating hair-cell adaptation into cochlear implant processing. *Hearing Research*, *342*, 48–57. https://doi.org/10.1016/j.heares.2016.09.008

Basilar membrane. (2022). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Basilar_membrane&oldid=1102122197

Beurg, M., Barlow, A., Furness, D. N., & Fettiplace, R. (2019). A Tmc1 mutation reduces calcium permeability and expression of mechanoelectrical transduction channels in cochlear hair cells. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(41), 20743–20749. https://doi.org/10.1073/pnas.1908058116

Beurg, M., Fettiplace, R., Nam, J.-H., & Ricci, A. J. (2009). Localization of inner hair cell mechanotransducer channels using high speed calcium imaging. *Nature Neuroscience*, *12*(5), 553–558. https://doi.org/10.1038/nn.2295

Beurg, M., Nam, J.-H., Crawford, A., & Fettiplace, R. (2008). The Actions of Calcium on Hair Bundle Mechanics in Mammalian Cochlear Hair Cells. *Biophysical Journal*, *94*(7), 2639–2653. https://doi.org/10.1529/biophysj.107.123257

Caprara, G. A., Mecca, A. A., & Peng, A. W. (2020). Decades-old model of slow adaptation in sensory hair cells is not supported in mammals. *Science Advances*, *6*(33), eabb4922. https://doi.org/10.1126/sciadv.abb4922

Corns, L. F., Johnson, S. L., Kros, C. J., & Marcotti, W. (2014). Calcium entry into stereocilia drives adaptation of the mechanoelectrical transducer current of mammalian cochlear hair cells. *Proceedings of the National Academy of Sciences*, *111*(41), 14918–14923. https://doi.org/10.1073/pnas.1409920111

Creel, D. J. (2015, June 18). *Figure 24. [Auditory brainstem response (ABR) recorded...]*. [Text]. University of Utah Health Sciences Center. https://www.ncbi.nlm.nih.gov/books/NBK303985/figure/CreelAlbinism.F24/

Einstein, A. (n.d.). *Investigations on the Theory of the Brownian Movement*. 11.

Fettiplace, R. (2017). Hair cell transduction, tuning and synaptic transmission in the mammalian cochlea. *Comprehensive Physiology*, *7*(4), 1197–1227. https://doi.org/10.1002/cphy.c160049

Gillespie, P. G. (2004). Myosin I and adaptation of mechanical transduction by the inner ear. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*(1452), 1945–1951. https://doi.org/10.1098/rstb.2004.1564

Gillespie, P. G., & Cyr, J. L. (2004). Myosin-1c, the hair cell's adaptation motor. *Annual Review of Physiology*, *66*, 521–545. https://doi.org/10.1146/annurev.physiol.66.032102.112842

Goutman, J. D., Elgoyhen, A. B., & Gómez-Casati, M. E. (2015). Cochlear hair cells: The sound-sensing machines. *FEBS Letters*, *589*(22), 3354–3361. https://doi.org/10.1016/j.febslet.2015.08.030

Khalighinejad, B., Herrero, J. L., Mehta, A. D., & Mesgarani, N. (2019). Adaptation of the human auditory cortex to changing background noise. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-10611-4

Li, S., Mecca, A., Kim, J., Caprara, G. A., Wagner, E. L., Du, T.-T., Petrov, L., Xu, W., Cui, R., Rebustini, I. T., Kachar, B., Peng, A. W., & Shin, J.-B. (2020). Myosin-VIIa is expressed in multiple isoforms and essential for tensioning the hair cell mechanotransduction complex. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-15936-z

Maoiléidigh, D. Ó., & Ricci, A. J. (2019). A bundle of mechanisms: Inner-ear hair-cell mechanotransduction. *Trends in Neurosciences*, *42*(3), 221–236. https://doi.org/10.1016/j.tins.2018.12.006

McPherson, D. R. (2018). Sensory Hair Cells: An Introduction to Structure and Physiology. *Integrative and Comparative Biology*, *58*(2), 282–300. https://doi.org/10.1093/icb/icy064

Myosin. (2022). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Myosin&oldid=1104292172

Peng, A. W., Effertz, T., & Ricci, A. J. (2013). Adaptation of mammalian auditory hair cell mechanotransduction is independent of calcium entry. *Neuron*, *80*(4), 960–972. https://doi.org/10.1016/j.neuron.2013.08.025

Perception Lecture Notes: Frequency Tuning and Pitch Perception. (n.d.). Retrieved August 29, 2022, from https://www.cns.nyu.edu/~david/courses/perception/lecturenotes/pitch/pitch.html

Pérez-González, D., & Malmierca, M. S. (2014). Adaptation in the auditory system: An overview. *Frontiers in Integrative Neuroscience*, *8*, 19. https://doi.org/10.3389/fnint.2014.00019

What is Myosin? (n.d.). MBInfo. Retrieved August 15, 2022, from https://www.mechanobio.info/cytoskeleton-dynamics/what-are-motor-proteins/what-is-myosin/

The Efficacy of Repetitive Transcranial Magnetic Stimulation (rTMS) in Stroke Rehabilitation of the Upper Extremities: A Scoping Review

Yuanyuan Xue

Author Background: *Yuanyuan Xue grew up in China and currently attends The Experimental High School Attached to Beijing Normal University in Beijing, China. Her Pioneer research concentration was in the field of biology/neuroscience and titled "Treatment of Arm Dysfunction After a Stroke."*

Abstract

Stroke is a leading cause of death and disability globally, and arm dysfunction is a common symptom in stroke patients. One of the interventions aiming to improve patients' upper limb functions post-stroke is Repetitive Transcranial Magnetic Stimulation (rTMS). rTMS is a non-invasive brain stimulation method that modulates cortical excitability, which may enhance motor recovery in stroke patients. Twenty randomized controlled trials (RCT) on rTMS and 8 RCTs on transcranial Direct Current Stimulation (tDCS) were included in this review. rTMS was found to be effective on function, motor function, muscle strength, spasticity, and quality of life. Moreover, correlations between cortical changes and motor recovery were reported. rTMS was compared with tDCS, which is cheaper and has a shorter duration than rTMS. rTMS was found to be more effective in muscle strength and quality of life than tDCS, while tDCS may have a potential effect on sensory functions.

1. Introduction

1.1. Stroke and Arm Dysfunction after Stroke

Stroke, caused by lack of blood supply or hemorrhage in the brain, is a principal cause of long-term serious disability and the global second leading cause of death.[1] Each year, approximately 795,000 people experience a new or reoccurred stroke.[2] There are two main categories of stroke—ischemic stroke (caused by blocking of blood vessels) and hemorrhagic stroke (caused by rupture of blood vessels, which leads to bleeding). Of the two types, ischemic stroke happens much more frequently, accounting for 87% percent of stroke events.[2]

One of the most common symptoms after stroke is arm dysfunction. Research has shown that 76% of survivors of a middle cerebral artery stroke

experience arm weakness and 62% of them do not achieve full dexterity of the upper limb at 6 months post-stroke.[3] Common problems in upper limbs after stroke include weakness, learned non-use, sensory loss, spasticity, and abnormal synergy. Therefore, studies aiming to examine the effect of treatments to treat dysfunction of the upper extremity post-stroke are meaningful to improving stroke patients' health conditions.[4–6]

Repetitive transcranial magnetic stimulation (rTMS) is believed to improve arm dysfunction after stroke through modulation of brain activities. After stroke, the impacted brain region experiences necroses (abnormal cell death) in the central area and a penumbra (a type of cell dysfunction) in the surrounding area. The brain will undergo a spontaneous reorganization that aims at repairing the impaired neural tissues in the brain, which often includes angiogenesis (proliferation of capillaries) and healing processes by inflammatory cells. Through the process of reorganization, the synaptic organization of the impacted area as well as that of distant areas is changed. Therefore, the balance between inhibitory and excitatory projections could be affected post-stroke, depending on the activity of surviving neurons.[7] As a result, the function of the upper extremities, which is controlled by the motor cortex of the brain, would also be affected by this change.

1.2. Overview of Repetitive Transcranial Magnetic Stimulation

Repetitive transcranial magnetic stimulation (rTMS) is a non-invasive and painless strategy that applies stimulation to the brain through currents generated by changing magnetic fields.[8] In stroke rehabilitation, rTMS is proposed to serve a role in regulating the excitability of the cerebral cortex and modulating the interhemispheric competition, which is often altered post-stroke.[9] There are two ways through which rTMS may enhance motor recovery: enhancing cortical excitability of the ipsilesional hemisphere or reducing cortical excitability of the contralesional hemisphere.[10] Studies have reported that high-frequency rTMS (HF-rTMS) induces long-term potentiation (LTP), while low-frequency rTMS (LF-rTMS) induces long-term depression (LTD).[8,11,12] LTP and LTD are two molecular-level mechanisms of neuroplasticity that reside in long-term changes of synaptic efficacy,[11] which produce long-lasting increase and decrease, respectively, of signal transmission between two neurons.[13] These two mechanisms are believed to play an important role in the learning and memory process,[11] and therefore could help with motor learning for stroke patients and explain the effectiveness of rTMS in stroke rehabilitation.

Theta burst stimulation (TBS) is another form of rTMS that is also found to modulate cortical excitability.[14] TBS delivers patterned stimulation pulses in bursts (triplets) at 50-Hz with an interval of 200ms, corresponding to the theta stimulation of the brain, which has been shown to be associated with learning and associative memory.[15–17] TBS also has a shorter duration as compared to conventional rTMS.[14,15] There are two subtypes of TBS: intermittent theta-burst stimulation (iTBS) and continuous theta-burst stimulation (cTBS). Similar to rTMS, the two types of TBS are facilitatory (iTBS) and inhibitory (cTBS), respectively. TBS shows effects in modulating brain excitability through LTP and LTD mechanisms as well.[16]

The motion of muscles is controlled by the brain motor cortex and their

excitability is closely related to brain excitability.[18] Therefore, the modulation of brain excitability by rTMS and TBS could help to increase function and activity of the paretic side while suppressing activities of the unaffected side, which could also help to prevent learned non-use. As a result, rTMS is believed to help improve brain plasticity and therefore facilitate the functional restoration of movement after stroke.[12]

1.3. Comparison of rTMS with Transcranial Direct Current Stimulation (tDCS)

Following a discussion of the evidence supporting the effects of TMS post-stroke, this review will also include a comparison between rTMS and tDCS. The similarities and differences between rTMS and tDCS will be discussed from different aspects, including features of the intervention, mechanism of effect, appropriate/inappropriate patients, as well as the effects as shown by clinical trials.

This scoping review includes 20 randomized controlled trials (RCTs) and 5 systematic reviews and meta-analyses that evaluated the effectiveness of repetitive transcranial stimulation (rTMS) of the upper extremity post-stroke. Additionally, 8 randomized controlled trials are reviewed to compare rTMS with tDCS.

2. Evidence Review

In this section, evidence of the immediate and sustained effects of two subtypes of repetitive transcranial magnetic stimulation (traditional rTMS and Theta Burst Stimulation) will be analyzed. This review will also summarize and discuss the adverse effects and appropriate patients for rTMS stimulation. Table 1 summarizes the classification of assessments used in the studies reviewed.

Table 1. *Assessment Classification*

	Classification	Assessments
1	function	Action Research Arm Test (ARAT)
2		Barthel Index (BI)
3		modified Barthel Index (mBI)
4		Korean-modified Barthel Index (K-mBI)
5		modified Rank Score (mRS)
6		Functional Independence Measure (FIM)
7		Finger Tapping Test (FTT)
8		Wolf Motor Function Test (WMFT)
9		Manual Function Test (MFT)
10		Motor Activity Log (MAL)
11		Box and Block Test (BBT)

	Classification	Assessments
12		Nine-Hole Peg Test (9HPT)
13		Jebsen-Taylor Hand Function Test (JTHFT)
14		Purdue Pegboard Test (PPT)
15	motor function	Fugl-Meyer Assessment (FMA)
16		Brunnstrom Recovery Stage (BRS)
17		Range of Motion (ROM)
18		National Institutes of Health Stroke Scale (NIHSS)
19	muscle strength	Medical Research Council (MRC)
20		Hand grip strength (HGS)
21		Motricity Index (MI)
22		Manual Muscle Test (MMT)
23		Preloading phase Duration (PLD)
24		Unloading phase Duration (ULD)
25		Grip force/Load force ratio (GFL/LFL)
26	spasticity	Modified Ashworth Scale (MAS)
27		Modified Tardieu Scale (MTS)
28		Shear Wave Ultrasound Elastography (SWV)
29	cortical excitability	Motor Evoked Potential (MEP) amplitude and latency
30		resting Motor Threshold (rMT)
31		Central Motor Conduction Time (CMCT)
32		Laterality Index (LI)
33		Motor Map Area
34		Functional Magnetic Resonance Imaging (fMRI) Activation
35	quality of life	Stroke Impact Scale (SIS)
36		Stroke-Specific Quality-of-Life scale (SSQOL)
37		Semmes Weinstein Monofilament Test (SWMT)

2.1. Effects of rTMS

In this section, the immediate and sustained effects of rTMS on function, motor function, muscle strength, spasticity, quality of life, and cortical excitability are discussed. The immediate effects are defined as the score changes between pre-to-post treatment sessions and sustained effects are those follow-up results. A total of 11 studies using repetitive transcranial magnetic stimulation are reviewed. Table 2 summarizes the information about studies evaluating the effects of rTMS on the upper limb post-stroke.

Table 2. *Summary of studies of rTMS*

Author, year study design sample size time post stroke category	Interventions Duration: session length, frequency per week for total number of weeks	Outcome measures Results (direction of effect)
Harvey et al., 2018 **RCT** **N_{start}=199** **N_{end}=169** **TPS=chronic**	E: 1Hz rTMS C: sham rTMS Duration: 3d/week, 6wks, 60-min physical therapy (15min of real/sham stimulation before)	• FMA: E & C (+post, ***), E vs C (-) • ARAT: E & C (+post, ***), E vs C (-) • WMFT: E & C (+post, ***), E vs C (-)
Du et al., 2016 **RCT** **N_{start}=69** **N_{end}=59** **TPS=acute & subacute**	E1: 1-Hz contralesional rTMS E2: 3-Hz ipsilesional rTMS C: sham rTMS Duration: 1h physical therapy, 5 consecutive days	• FMA: E1 vs C (+exp, *), E2 vs C (-), E1 vs E2 (-) • MRC: E1 vs C (+exp, **) at 2 months, and (+exp, ***) at 3 months; E2 vs C (-); E1 vs E2 (-) • BI: E1 vs C (+exp, ***), E2 vs C (+exp, **), E1 vs E2 (-) • mRS: E1 vs C (+exp, *), E2 vs C (+exp, *) at EOT, 2 months and 3 months; E1 vs E2 (-) • rMT, MEP: increase in AH: E1 vs C (+exp, ***), E2 vs C (+exp, ***), E1 vs E2 (-) suppression on UH: E1 (+post, *), E2 and C (-); between group: (-)
Guan et al., 2017 **RCT** **N_{start}=42** **N_{end}=27** **TPS=acute**	E: 5-Hz rTMS C: sham rTMS Duration: 10 consecutive days	• NIHSS: E vs C (+exp, *) at 2 days, 1 month; (-) at 3 and 6 months, 1 year • FMA: E vs C (+exp, *) at 2 days, 1, 3, and 6 months, 1 year • BI: E vs C (+exp, *) at 2 days, 1 month; (-) at 3 and 6 months, 1 year • mRS and MT: E vs C (-)

Author, year study design sample size time post stroke category	Interventions Duration: session length, frequency per week for total number of weeks	Outcome measures Results (direction of effect)
Kim et al., 2020 **RCT** **N_{start}=77** **N_{end}=73** **TPS=subacute**	E: 1-Hz rTMS C: sham rTMS Duration: 30min/d + 30min task-based occupational therapy, 10 consecutive days	E vs C: overall (-) subgroup with cortical involvement: (-) subgroup with no cortical involvement: • BBT: E vs C (+exp, *) for FAS analysis; (+exp, **) for PP analysis • FTT: E vs C (+exp, *) for PP analysis at 1 month; (-) for FAS analysis • BRS: E vs C (+exp, *) for FAS analysis, (+exp, **) at for PP analysis at EOT • HGS: E vs C (+exp, *) for PP analysis, (-) for Bonferroni correction
Du et al., 2019 **RCT** **N_{start}=60** **N_{end}=44** **TPS=acute**	E1: 10 Hz rTMS E2: 1 Hz rTMS C: sham rTMS Duration: 20~26min/d, 5 consecutive days	• FMA: E1, E2, E3, C (+post, ***); E1 vs C (+exp, *), E2 vs C (+exp, *), E1 vs E2 (-) • rMT, CMCT: rMT & CMCT (AH): E1, E2 (+post, **); C (-) CMCT (UH): E1 (+post, **); E2, C (-) between group: (-) • fMRI activation: ipsilesional M1: E1 vs C (-), E2 vs C (-), E1 vs E2 (+exp1, **) contralesional M1: E1 vs C (-), E2 vs C (+exp, *), E1 vs E2 (+exp2, *) SMA: E1 vs C (-), E2 vs C (-), E1 vs E2 (+exp1, **)
Sasaki et al., 2013 **RCT** **N_{start}=29** **N_{end}=29** **TPS=acute**	E1: 10-Hz rTMS E2: 1-Hz rTMS C: sham rTMS Duration: 45min/d, 5 consecutive days, together with conventional rehabilitation	• HGS: E1 & E2 (+post, *), C (-); E1 vs C (+exp, *); E2 vs C (-); E1 vs E2 (-) • FTT: E1 & E2 (+post, *), C (-); E1 vs C (+exp, *); E2 vs C (-); E1 vs E2 (-)

Author, year study design sample size time post stroke category	Interventions Duration: session length, frequency per week for total number of weeks	Outcome measures Results (direction of effect)
Yang et al., 2021 **CT** **N_{start}=44** **N_{end}=39** **TPS=subacute**	E1: 5-Hz rTMS + hand grip training E2: 5-Hz rTMS E3: hand grip training Duration: 5min/d, 5d/wk, 2wks, together with 120min physical therapy	• JTHFT: E1,E2,E3 (+post, **) E1 vs E3 (+exp1, *), E1 vs E2 (+exp1, *), E1 vs E2 (-) • FMA: E1,E2,E3 (+post, **) E1 vs E3 (+post, *), E1 vs E2 (+, **), E1 vs E2 (-) • HGS: E1 (+post, ***), E2 & E3 (-) between group: (-) • mBI: E1,E2,E3 (+post, *); between group: (-) • iMEP latency: (-)
Hosomi et al., 2016 **RCT** **N_{start}=41** **N_{end}=39** **TPS=subacute**	E: 5-Hz rTMS on ipsilesional M1 C: sham rTMS Duration: 10min/d, 5d/wk, 2wks	• BS: E vs C: hand score (+exp, *), arm (-), lower limb (-) at 17 days • FMA: E & C (+post, *) for the total score, proximal upper limb score at EOT at 17 days; E (+), C (-) at EOT; E & C (+post, *) for distal upper limb score at 17 days E vs C (-) • NIHSS: E & C (+post, *) for total score; E (+post, *), C (-) for motor arm score at EOT and at 17 days; E vs C (-) • FIM: E & C (+post, *) for subscores; E vs C (-) • FTT: within group: E (+post, **), C (-); E vs C (-) • HGS: within group: E (+post, *) at 17 days, C (-); E vs C (-)

Author, year study design sample size time post stroke category	Interventions Duration: session length, frequency per week for total number of weeks	Outcome measures Results (direction of effect)
Sharma et al., 2020 **RCT** **N_{start}=96** **N_{end}=89** **TPS=subacute**	E: 1-Hz rTMS on contralesional M1 C: sham rTMS on contralesional M1 Duration: 5d/wk, 2wks, together with 45-50 min physiotherapy	• mBI: E vs C (+exp, ***) • FMA: E (+post, ***); E vs C (-) • NIHSS: E vs C (+exp, ***) • mRS: E (+post, ***), C (-); E vs C (-)
Galvão et al., 2014 **RCT** **N_{start}=20** **N_{end}=18** **TPS=chronic**	E: 1-Hz rTMS to M1 of UH C: sham rTMS to M1 of UH Duration: 3d/week, 10 sessions in total, together with physical therapy	• MAS: E (+post, ***) at EOT, (p=0.03, -) at 1 month, C (-); E vs C (+exp, ***) • FMA: E & C (+post, *) at EOT & 1 month; E vs C (-) • FIM: E (+post, *), C (-); E vs C (-) • wrist ROM: E (+post, *), C (-); E vs C (-) • SSQOL: E & C (+post, *) at EOT, (-) at 1 month; E vs C (-)
Pinto et al., 2019 **RCT** **N_{start}=27** **N_{end}=27** **TPS=chronic**	E1: 1-Hz rTMS + fluoxetine E2: sham rTMS + fluoxetine C: sham rTMS + placebo fluoxetine Duration: 5d/wk*2wks + 1d/wk*8wks, 20min/d, total 18 sessions over 90d	• JTHFT: E1 vs E2 (+exp1, ***), E1 vs C (+exp, **), E2 vs C (+con, *) • FMA: E2 vs C (+con, *) • MAS: (-) • MEP amplitude: E1 vs C (+exp, *) • ICI, ICF: E1 vs C (+exp, *); ICF (-)

Abbreviations and table notes: C=control group; E=experimental group; h=hours; min=minutes; d=days; wks=weeks; RCT=randomized controlled trial; TPS=time post stroke category (acute: less than 1 month, subacute: more than 1 month but less than 6 months, chronic: more than 6 months); UH=unaffected hemisphere; AH=affected hemisphere; EOT=end of treatment; PP: per protocol; FAS: full analysis set

"+post" indicates a significant within-group difference in favor of post-treatment scores
"+pre" indicates a significant within-group difference in favor of baseline scores
"+exp" indicates a significant between-group difference in favor of the experimental group
"+exp1" indicates a significant between-group difference in favor of the first experimental group
"+exp2" indicates a significant between-group difference in favor of the second experimental group
"+con" indicates a significant between-group difference in favor of the control group
"-" indicates no significant within-group/between-group difference
"*" indicates a significance level of $p<0.05$; "**" indicates a significance level of $p<0.01$; "***" indicates a significance level of $p<0.001$.
"at xx months/wks" means that the result is tested at xx months or weeks post-intervention

2.1.1. Effect of rTMS on Function

Functional tests of stroke recovery focus on patients' level of independence.[19] After a stroke, many patients are unable to independently complete daily tasks. Forty-six percent of stroke patients are, to different degrees, functionally dependent on others.[20] Function is typically examined by evaluating patients' performance in various activities of daily living (ADL), hand dexterity, or manual skills. In the studies reviewed, 10[21–30] of the 11 trials used functional assessments and 7 studies utilized multiple tests. The outcome measures used include the Action Research Arm Test (ARAT), Barthel Index (BI), modified Barthel Index (mBI), modified Rank Score (mRS), Functional Independence Measure (FIM), Finger Tapping Test (FTT), Wolf Motor Function Test (WMFT), Box and Block Test (BBT), and Jebsen-Taylor Hand Function Test (JTHFT).

2.1.1.1. Immediate Effect of rTMS on Function

All 10 studies reviewed reported significant between-group or within-group differences immediately after treatment. Seven studies[22–26,28,30] reported statistically significant differences between experimental and control groups favoring the experimental group. Two of these 7 studies used 2 experimental groups—an HF-rTMS group and an LF-rTMS group—to compare with the control group.[22,25] In both studies, statistically significant differences were reported favoring the LF-rTMS and the HF-rTMS groups over controls. In one study,[26] 3 experimental groups were used (rTMS + hand grip training; rTMS only; hand grip training only). The study reported significantly better results in the combined group compared to the hand-grip-training-only group.[26] In addition, in the 3 studies that reported no between-group differences, 2 studies[27,29] reported significant within-group differences in experimental groups only on FTT and FIM tests, respectively; however, Harvey et al.[21] reported significant within-group differences on ARAT and WMFT in both experimental and control groups. Notably, in one study,[30] significant differences were reported between the combined group (fluoxetine + rTMS) and the fluoxetine group (fluoxetine + sham), and between combined and control groups, both favoring the combined group. In this study, the fluoxetine group's pre-to-post intervention improvements were less than that of the control group. The authors suggest that this may be due to the negative effect of fluoxetine and that the positive effect of rTMS may have counteracted this negative effect.

In studies where multiple assessments were used, there are differences in significance by different assessments. In the 7 studies[21–24,26–28] that used more than one assessment, 4[23,26–28] reported different outcomes in different assessments. Guan and colleagues reported significant between-group differences favoring the experimental group in mBI (measures general functions), while no significance in mRS (measures level of independence).[23] The study by Sharma et al. also reported between-group differences in mBI but not in mRS.[28] This might be due to the characteristics of mRS. Since mRS is a subjectively assigned scale and studies had also mentioned its lack of clear criteria and low reliability, this ambiguity may explain the lack of significance in the study.[31–33] In addition, Hosomi et al. used FIM and FTT, and Yang et al. used JTHFT and mBI in their studies.[26,27] While FIM and mBI are composite measures of overall functions, JTHFT and FTT focus more on hand functions. Therefore, the differences in significance may also be due to the different aspects

that different assessments measure. As a result, the lack of significance in some assessments of a study might be because of the psychometric features of the assessment or the aspect it assesses.

In conclusion, in the 10 studies examining function, 7 reported significant between-group differences favoring rTMS, 2 studies reported significant within-group changes in the experimental group only, and 1 study reported significant changes in both experimental and control groups. These results suggest that rTMS does have a positive effect on improving functional abilities of the upper extremities post-stroke.

2.1.1.2. Sustained Effect of rTMS on Function

In the 5 studies that designed follow-ups, 3 studies reported between-group significance at follow-ups. Du et al. reported significant between-group differences in mRS between the experimental and control groups favoring the 2 experimental groups at both 2 and 3 months after the intervention. In addition, Guan et al. and Kim et al. reported significant between-group differences favoring the experimental groups at 1-month post-intervention, in BI and FTT, respectively. This suggests that the effect of rTMS on function could persist after the intervention.

2.1.2. Effect of rTMS on Motor Function

Motor function is an important index that is frequently assessed post-stroke. It may be indicative of the severity of the stroke and the level of recovery after the stroke. In the RCTs reviewed, 10 of the 11 studies examined the effect of rTMS on motor function. The outcome measures used to assess motor function of the upper limb include the Fugl-Meyer Assessment (FMA), Brunnstrom Recovery Stage (BRS), Range of Motion (ROM), and National Institutes of Health Stroke Scale (NIHSS).

2.1.2.1. Immediate Effect of rTMS on Motor Function

Ten studies reported significant between-group or within-group differences in motor function immediately after the intervention. Six[22–24,26,28,34] of the 10 studies reviewed that examined motor function reported statistically significant differences between the experimental group (rTMS) and control (sham stimulation) groups post-treatment. Among the five, two[22,34] had multiple experimental groups (LF-rTMS and HF-rTMS groups). In one study,[34] both experimental groups (10-Hz rTMS and 1-Hz rTMS groups) produced significantly better results compared to the control group. In the other study,[22] the group using 1-Hz rTMS had a significant between-group difference (LF-rTMS vs. control) favoring the 1-Hz rTMS group, while the experimental group using 3-Hz rTMS reported a lack of significance. In addition, in studies where no between-group (experimental vs. control) difference was found, 2 studies reported significant within-group differences in experimental groups only.[27,29] Specifically, the differences were reported in the NIHSS motor arm score and wrist ROM in the 2 studies, respectively. Hosomi and colleagues[27] also noted earlier improvement in the FMA distal upper limb subscore and NIHSS total score for the experimental group. In addition, 2 studies reported significant within-group differences in both the experimental group and the control group.[21,30]

There are possible reasons for the lack of significant differences between the control and experimental groups. Though Pinto and colleagues reported no significant differences between the combination group (rTMS + fluoxetine, fluoxetine as adjuvant therapy) and the fluoxetine group (fluoxetine only), or between the combination group and the control group, a significant difference favoring the control was found in the comparison between the fluoxetine group and the control group.[30] This may indicate that fluoxetine has a negative effect on motor function, so the lack of significance in the combined vs. control group may be because rTMS counteracted this negative effect. Another reason might be the characteristics of FMA, which is the most commonly used assessment of motor function in the studies reviewed. The study by Harvey et al. that reported within-group significance only and in both experimental and control groups also used FMA.[21] Several studies[30,35–37] had found the responsiveness of FMA to be small or moderate; the relatively low sensitivity of the FMA is mentioned by Pinto and colleagues as well.[30] In summary, the lack of significance in some studies might be due to the treatment protocol of the study or the psychometric properties of the assessments used.

To sum up, in the 10 studies reviewed in this section, 6 studies reported significant between-group differences in motor function favoring rTMS groups,[22–24,26,28,34] 2 studies reported significant within-group changes in the experimental group only,[27,29] and 2 studies reported significant within-group changes in both experimental and control groups.[21,30] This result suggests that rTMS has an overall positive effect on motor function of upper extremities post-stroke, but given the strength of evidence further studies are warranted. Currently, FMA is used most frequently (all 10 studies used FMA) to examine motor function, while fewer studies are using NIHSS, BRS, ROM and other motor function tests.[23,24,27–29] This may suggest that FMA is an assessment that receives the most recognition from researchers. As a result, future studies studying the effect of rTMS on motor function may want to use the FMA measure so that they can be more comparable to other studies.

2.1.2.2. Sustained Effect of rTMS on Motor Function

Three studies reported significant between-group or within-group differences in the 4 studies that had follow-ups. Guan et al. reported significant between-group differences favoring the experimental group in FMA at 1, 3, 6 months, and 1-year post-intervention and in NIHSS at 1-month post-intervention. Hosomi et al. reported significant within-group improvements in NIHSS motor arm score from baseline to 17 days post-intervention in the experimental group only. Significant within-group differences in both experimental and control groups were also reported by Galvão at 1-month post-intervention and by Hosomi et al. at 17 days post-intervention in FMA. These significant results suggest that there is a sustained effect of rTMS on motor function. However, the evidence is less robust when compared to the function section. As a result, future studies could further investigate patients' motor function at follow-ups.

2.1.3. Effect of rTMS on Muscle Strength

Muscle strength is another common construct measured post-stroke. Five of the 11 studies reviewed in this subsection examined muscle strength following rTMS, and 7 outcomes including Hand Grip Strength (HGS)[24–27] and the

Medical Research Council Scale (MRCS)[22]. The HGS includes 1 functional task that focuses on the strength of the hand muscles,[38] while MRCS is an assessment evaluating gross muscle strength.[39]

2.1.3.1. Immediate Effect of rTMS on Muscle Strength

Three studies reported significant between-group or within-group differences immediately after treatment.[24–26] Significant between-group (rTMS vs. sham) differences were reported in 2 studies,[24,25] suggesting that rTMS may be effective in improving muscle strength. Specifically, Sasaki et al.[25] compared two experimental groups using HF-rTMS or LF-rTMS and reported a significant between-group difference favoring rTMS between the HF-rTMS and control groups only. Significant within-group differences before and after the treatment were reported in 2 studies[25,26] on the HGS assessment. Both studies reported significant within-group pre-to-post differences in the experimental groups only.[26] Yang et al. used 3 experimental groups in total (rTMS + hand grip training; rTMS only; hand grip training only). In this trial, significant within-group improvement was reported only in the combined group.[26] Sasaki et al. also reported significant improvement from baseline to post-intervention in LF-rTMS and HF-rTMS groups only.[25] However, Du et al.[22] and Hosomi et al.[27] reported no significant differences immediately after treatment.

In conclusion, in the 4 studies[22,24–26] reviewed in this section, 2 studies[24,25] reported significant between-group differences favoring rTMS, 1 study[26] reported significant within-group pre-post changes in the experimental group only, and 1 study[22] reported no significance immediately after the treatment. These mixed results suggest that rTMS may have potential effects on muscle strength of upper extremities for stroke rehabilitation, but the evidence is inconclusive, so more studies are needed to further investigate the effectiveness of rTMS on muscle strength. Three studies reviewed in this section used HGS (measures hand muscle strength), in contrast with only 1 study using MRC (measures gross muscle strength), which may suggest that there are fewer studies on the gross muscle strength of the upper limbs. As a result, future studies could use MRC more frequently to provide more evidence for the effect of rTMS on muscle strength of the upper limb other than grip force of the hand.

2.1.3.2. Sustained Effect of rTMS on Muscle Strength

Two studies conducted follow-ups after the intervention was over. Du et al. reported statistically significant differences in MRCS scores between the LF-rTMS group and the control group at 2 months and 3 months. Moreover, Hosomi et al. reported significant within-group differences from baseline to 17 days after treatment in the experimental group only. These results suggest that the effect of rTMS on muscle strength could be sustained to the checkpoints set by each study.

2.1.4. Effect of rTMS on Spasticity

Spasticity is a common symptom after stroke, with 30% of patients having this disorder.[40] Spasticity severely affects upper limb functions. It may also influence the quality of life and limit stroke recovery.[4,41,42] One study reported that severe spasticity more frequently occurs in the upper limbs compared to the lower limbs.[43] Therefore, studying the effect of rTMS on spasticity is important in

stroke rehabilitation. Both studies[29,30] reviewed in this section used the Modified Ashworth scale (MAS) to quantify spasticity.

2.1.4.1. Immediate Effect of rTMS on Spasticity

Galvão et al. reported a statistically significant difference favoring the rTMS versus the control group on the MAS.[29] Moreover, the study reported that, immediately after the intervention, 90% of patients in the experimental group showed clinically important improvement in the MAS score, while only 30% of patients in the control group showed clinically important improvement.[29] However, Pinto et al. reported no significant differences between rTMS and control groups.[30] No significance was reported at follow-ups. In conclusion, only one of the two reviewed studies reported significant between-group differences in spasticity, favoring rTMS. As a result, the evidence regarding the effect of rTMS on spasticity reviewed in this study is mixed. More evidence is needed to provide a valid conclusion on this topic. The 2 studies reviewed used sample sizes of 20 and 27, respectively, which are not large sample sizes. Both studies mentioned the limitation of their small sample sizes and one also noted the heterogenous sample might have also affected the detection of effects. Therefore, future studies using larger sample sizes may detect the effects of rTMS better and provide more evidence.

2.1.5. Effect of rTMS on Quality of Life

After a stroke, quality of life (QOL) is often affected. One study[44] showed that 43% of mildly affected stroke survivors experienced degradation in QOL compared to their pre-stroke conditions. One RCT using rTMS intervention that examined patients' improvement in QOL was reviewed.[29] The assessment used is the Stroke-Specific Quality-of-Life scale (SSQOL), which is a self-reported scale that includes 12 domains and 49 items in total.[45] In this RCT, both experimental and the control groups showed significant improvement in SSQOL scores from baseline to post-intervention and no between-group difference was found. Also, no significant results were reported at follow-up. This result suggests that using rTMS may not be effective to improve QOL post-stroke. However, this evidence is very preliminary, since it is based on only one study with a small sample size (n=20).[29] Future studies could set larger sample sizes and use SSQOL as the outcome measure so that results across different studies can be compared. Since there is only 1 RCT reviewed, SSQOL may not be the most common assessment used; as a result, future studies using different measures of QOL will also provide valuable evidence and may help to find the most appropriate outcome measures of QOL.[29]

2.1.6. Effect of rTMS on Cortical Excitability

It has been hypothesized that rTMS helps improve post-stroke recovery by modulating the interhemispheric balance of the brain[18]. As a result, cortical excitability has been used as an outcome in multiple studies. Of the 11 studies reviewed, 5 studies[22,23,26,30,34] measured cortical excitability, using indices including resting motor threshold (rMT), motor evoked potential (MEP) amplitude and latency, Central Motor Conduction Time (CMCT), functional magnetic resonance imaging (fMRI) activation, and Intracortical Inhibition

(ICI). No significant results were reported at follow-ups.

2.1.6.1. Immediate Effect of rTMS on Cortical Excitability

Three studies[22,30,34] reported statistically significant differences between groups and 2 studies[23,26] reported no significance. In detail, Du et al. reported significant differences in increased excitability of the affected hemisphere (AH) (indicated by reduced rMT, increased MEP amplitude and reduced MEP latency) in both experimental groups when compared to the control.[22] Du et al. provided similar evidence by reporting significant between-group differences in an increase of fMRI activation of ipsilesional M1 and supplementary motor area (SMA) favoring the HF-rTMS group compared to the LF-rTMS and control groups; significant between-group differences in the decrease of contralesional M1 fMRI activation favoring the LF-rTMS group compared to HF-rTMS and control groups were also reported.[34] Pinto et al. reported a significant difference between the combined therapy group (rTMS + fluoxetine) and the control group in MEP amplitude of the unaffected hemisphere (UH) and ICI favoring the combined group.[30] Since differences between fluoxetine-only and control groups were not reported, it can be assumed that this greater cortical excitability change is the effect of rTMS. Moreover, Du et al.[22] reported significantly decreased UH activity (indicated by decreased MEP amplitude and prolonged MEP latency) only in the 1-Hz rTMS group but not the 3-Hz rTMS group and the control group. This result is in accordance with the hypothesis that LF-rTMS may modulate cortical excitability by lowering the excitability of the unaffected side.[46–48]

In summary, in the 5 studies discussed in this part, 3 studies[22,30,34] reported significant between-group differences, and 2 studies[23,26] reported no significance immediately post-intervention. Therefore, the results of cortical excitability after rTMS are mixed. While rTMS shows a tendency toward improving interhemispheric balance in stroke patients by lowering the excitability of the UH or increasing the excitability of the AH, the insignificance in 2 of the studies suggests that further investigations are still needed to confirm this assumption. In the studies reviewed, the rMT and MEP are measured most frequently. Therefore, future studies examining the change of cortical excitability following rTMS could use these 2 indices.

Correlations between changes in cortical excitability and patients' motor function were analyzed in 2 studies. Du et al. reported a significant and positive correlation ($r=0.615$, $p<0.001$) between changes in rMT (from baseline to 3 months post-intervention) and upper limb score of FMA; in a later study, Du and colleagues also reported significant and positive correlations between ipsilesional M1 activation and the FMA score both immediately after intervention ($r=0.315$, $p=0.042$) and at 3 months post-intervention ($r=0.338$, $p=0.047$). These results support the assumption that rTMS improves stroke patients' motor performance by modulating cortical excitability.

2.2. Effect of TBS

Theta Burst Stimulation (TBS) is a form of repetitive transcranial magnetic stimulation (rTMS). It utilizes lower intensity stimulus pulses and has a shorter stimulation time,[14] believed to be more comfortable for patients.[49] Studies have

shown that TBS may produce equal or better effects than traditional rTMS on various aspects.[50] There are 9 studies[49,51–58] reviewed in this section. In this section, the immediate and sustained effects of TBS on function, motor function, spasticity, muscle strength, quality of life (QOL), and cortical excitability will be discussed. Table 3 summarizes the information about studies evaluating the effects of TBS on the upper limb post-stroke.

Table 3. *Summary of studies of TBS*

Author, year study design sample size time post stroke category	Interventions Duration: session length, frequency per week for total number of weeks	Outcome measures Results (direction of effect)
Kuzu et al., 2021 **RCT** **N_{start}=20** **N_{end}=20** **TPS=chronic**	E1: 1-Hz rTMS on contralesional hemispheric UE M1 E2: cTBS C: sham cTBS Duration: 20min/d + 60min physical therapy, 10 sessions in total	• FMA: E1 & E2 (+post, *) at EOT and at 4 weeks; C (-) E1 vs C (+exp, **), E2 vs C (+exp, *) at EOT and at 4 weeks, E1 vs E2 (-) • MAS: elbow flexor: E1 (+post, *) at EOT and at 4 weeks, E2 (+post, *) at EOT, C (-) between group: (-) pronator: E1 (+post, *) at EOT and at 4 weeks, E2 & C (-) between group: (-) wrist flexor muscle groups: E1 & E2 (+post, *) at EOT and at 4 weeks; E2 vs C (+exp, *), E1 vs C (-), E1 vs E2 (-) finger flexor muscle groups: E1 (+post, *) at EOT and 4 weeks, E2 & C (-) between group: (-) • FIM & MAL-28: E1 & E2 (+post, *), C (-) at EOT
Chen et al., 2019 **RCT** **N_{start}=23** **N_{end}=22** **TPS=chronic**	E: iTBS on ipsilesional M1 C: sham iTBS on ipsilesional M1 Duration: about 20min/d, 5d/wk, 2wks, 90min conventional physical therapy	• MAS: E vs C (+exp, *) & larger effect sizes • FMA: E vs C (+exp, *) & larger effect sizes • ARAT: E vs C (+exp, ***) & larger effect sizes • BBT: E vs C (+exp, *) & larger effect sizes • MAL: E vs C (-)

Author, year study design sample size time post stroke category	Interventions Duration: session length, frequency per week for total number of weeks	Outcome measures Results (direction of effect)
Sung et al., 2013 **RCT** **N_{start}=54** **N_{end}=54** **TPS=chronic**	E1: 1-Hz rTMS over the contralesional M1 + iTBS over the ipsilesional M1 E2: sham rTMS + iTBS over the ipsilesional M1 E3: 1-Hz rTMS over the contralesional M1 + sham iTBS C: sham rTMS + sham iTBS Duration: 45min/d, 5d/wk, 4wks	MRC, FMA, WMFT, FTT, and RT: E1/E2/E3 vs C (+exp, *) • WMFT: E1 vs E2 (+exp1, ***), E1 vs E3 (+exp1, ***), E2 vs E3 (-) • FMA: E1 vs E2 (+exp1, ***), E1 vs E3 (+exp1, ***), E2 vs E3 (+exp3, *) • MRC: E1 vs E2 (+exp1, **), E1 vs E3 (+exp1, ***), E2 vs E3 (+exp3, *) • RT: E1 vs E2 (+exp1, ***), E1 vs E3 (+exp1, **), E2 vs E3 (-) • motor map area: contralesional: E1/E3 vs C (+exp, **), E2 vs C (+exp, *) ipsilesional: E1/E2 vs C (+exp, *) • MEP amplitude and latency: E1 vs C (+exp1, *)
Talelli et al., 2012 **semi-RCT** **N_{start}=41** **N_{end}=41** **TPS=chronic**	E1: iTBS over ipsilesional side C1: sham iTBS E2: cTBS over contralesional side C2: sham cTBS Duration: 60min/d, 5d/wk, 2wks, PT after each session	• 9HPT: E1 & C1 (+post, ***) at EOT, and (+post, **) at 30 days; E1 vs C1 (-) E2 & C2 (+post, **) at EOT and at 30 days, (+post, *) at 90 days; E2 vs C2 (-) • JTHFT: E1, C1, E2, C2 (+post, *) • HGS: pinch grip: (-) grasp: E1, C1, E2, C2 (+post, *)

Author, year study design sample size time post stroke category	Interventions Duration: session length, frequency per week for total number of weeks	Outcome measures Results (direction of effect)
Chen et al., 2021 **RCT** **N_{start}=32** **N_{end}=29** **TPS=subacute & chronic**	E: cerebellar iTBS on ipsilesional lateral cerebellum C: sham iTBS on ipsilesional lateral cerebellum Duration: 5d/wk, 2wks, 50min conventional physical therapy	• MAS: elbow flexors: E (+post, ***), C (+post, *); E vs C (exp+, **) wrist flexor: E (+post, ***), C (-); E vs C (+exp, **) • MTS: elbow flexors: E (+post, ***), C (+post, **); E vs C (+exp, ***) wrist flexors: E & C (+post, ***); E vs C (+exp, ***) • SWV: biceps brachii: E (+post, ***), C (+post, **); E vs C (+exp, *) flexor carpi radialis: E (+post, ***), C (+post, *); E vs C (+exp, ***) • BI: E & C (+post, ***); E vs C (-) • MEP: E (+post, **), C (-); E vs C (-) • H_{max}/M_{max} Ratio: C (+post, **), E (-); E vs C (-)

Author, year study design sample size time post stroke category	Interventions Duration: session length, frequency per week for total number of weeks	Outcome measures Results (direction of effect)
Meng et al., 2020 **RCT** **N_{start}=28** **N_{end}=28** **TPS=subacute**	E1: 1-Hz contralesional rTMS + iTBS on the ipsilesional side E2: 1-Hz contralesional rTMS + sham iTBS on the ipsilesional side E3: sham contralesional rTMS + iTBS on the ipsilesional side Duration: 5d/wk, 2wks, 1h conventional rehabilitation	• FMA: E1 (+post, **), E2 (+post, **), E3 (+post, *) E1 vs E2 (+exp1, *), E1 vs E3 (+exp1, **) • BI: E1 (+post, **), E2 (+post, **), E3 (+post, *) E1 vs E2 (+exp1, *), E1 vs E3 (+exp1, ***) • MEP: amplitude: EDC: E1 vs E2 (+exp1, *), E1 vs E3 (+exp1, ***) biceps brachii: E1 vs E2 (+exp1, **), E1 vs E3 (+exp1, **) APB: E1 vs E3 (+exp1, ***) latency: EDC: E1 vs E2 (+exp1, *), E1 vs E3 (+exp1, **) biceps brachii: E1 vs E2 (+exp1, **), E1 vs E3 (+exp1, ***) APB: E1 vs E3 (+exp1, ***)
Hsu et al., 2013 **RCT** **N_{start}=12** **N_{end}=12** **TPS=acute**	E: iTBS (1200 pulses, prolonged) C: sham stimulation Duration: 10min/d, 10 consecutive days, with conventional rehabilitation (2 sessions/d, 90min each)	• safe and well-tolerated • NIHSS: E & C (+post, *) at EOT and post-stroke day 60 (4-6 wks after EOT); E vs C (+exp, *) at EOT, (+exp, **) at post-stroke day 60 • FMA: E & C (+post, *) at EOT and post-stroke day 60; E & C (+post, *) for all 4 subtests E vs C (+exp, **) at EOT and post-stroke day 60 • ARAT: E & C (+post, *) at EOT and at post-stroke day 60; E vs C (-) • AH aMT and MEP: E vs C (-)

Author, year study design sample size time post stroke category	Interventions Duration: session length, frequency per week for total number of weeks	Outcome measures Results (direction of effect)
Chen et al., 2021 **RCT** **N_{start}=24** **N_{end}=23** **TPS=chronic**	E: iTBS stimulation C: sham stimulation Duration: ~5min/session, 2 sessions/d, 10min interval, 5d/wk, 3wks, with 60min VCT	• FMA: E (+post, *), C (+post, **) E vs C (-); 2 in the E group and 3 in the C group reached MCID • MAS: E (+post, **), C (-); E vs C (+exp, **); 7 in the E group and 1 in the C group reached MCID • ARAT: E (+post, *), C (+post, *); E vs C (+exp, *); 4 in the E group and 1 in the C group reached MCID • BBT: E (+post, *), C (-); E vs C (-) • 9HPT: E (+post, *), C (-); E vs C (+exp, *) • MAL: AOU: E (+post, **), C (-); E vs C (+exp, **) QOM: E (+post, **), C (-); E vs C (-) • SIS: E (+post, ***), C (-) E vs C (+exp, **); 2 in the E group and 0 in the C group reached MCID
Ackerley et al., 2016 **RCT** **N_{start}=18** **N_{end}=18** **TPS=chronic**	E: iTBS over ipsilesional M1 C: sham iTBS over ipsilesional M1 Duration: 5d/wk, 2wks, with 45min UL physical therapy	• ARAT: E (+post, **), C (-) at EOT; E (+post, **), C (-) at 1 month; E vs C (+exp, **) at EOT, and (+exp, *) at 1 month • FMA: E vs C (-) • LI: E vs C (-)

Abbreviations and table notes: C=control group; E=experimental group; h=hours; min=minutes; d=days; wks=weeks; RCT=randomized controlled trial; TPS=time post stroke category (acute: less than 1 month, subacute: more than 1 month but less than 6 months, chronic: more than 6 months); UH=unaffected hemisphere; AH=affected hemisphere; EOT=end of treatment; M1=primary motor cortex; RT=reaction time; UL=upper limb; VCT: virtual reality-based cycling training; MCID: minimal clinically important difference; AOU: amount of use; QOM: quality of movement

"+post" indicates a significant within-group difference in favor of post-treatment scores
"+pre" indicates a significant within-group difference in favor of baseline scores
"+exp" indicates a significant between-group difference in favor of the experimental group
"+exp1" indicates a significant between-group difference in favor of the first experimental group
"+exp3" indicates a significant between-group difference in favor of the third experimental group
"+con" indicates a significant between-group difference in favor of the control group
"-" indicates no significant within-group/between-group difference
"*" indicates a significance level of $p<0.05$; "**" indicates a significance level of $p<0.01$; "***" indicates a significance level of $p<0.001$.
"at xx months/wks" means that the result is tested at xx months or weeks post-intervention

2.2.1. Effect of TBS on Function

In the studies reviewed that used TBS, functional tests were also widely applied. All 9 studies[49,51–58] in the TBS section used assessments of functions and 4[49,51,53,55] of them used multiple tests. The assessments include the Action Research Arm Test (ARAT), Barthel Index (BI), modified Rank Score (mRS), Functional Independence Measure (FIM), Finger Tapping Test (FTT), Wolf Motor Function Test (WMFT), Motor Activity Log (MAL), Box and Block Test (BBT), and Jebsen-Taylor Hand Function Test (JTHFT).

2.2.1.1. Immediate Effect of TBS on Function

All 9 studies produced significant between-group or within-group differences immediately after the TBS intervention. Five studies[51,52,55,56,58] reported significant differences between the experimental and control groups favoring the experimental group. Among these 5 studies, Sung et al.[52] set up 3 experimental groups (1-Hz rTMS + iTBS; iTBS only; rTMS only) and a control group. All 3 experimental groups produced significantly better functional outcomes than the control group in WMFT (measures overall function) and FTT (measures hand function). Significant between-group differences favoring the combined group were also reported in WMFT between the combined group and the 2 other experimental groups (iTBS only; rTMS only). Two studies[51,55] also reported significant between-group differences favoring the experimental group in multiple tests. Yu-Jen Chen and colleagues reported significant between-group differences in both the ARAT (measures gross function) and the BBT (gross manual dexterity);[51] Yu-Hsin Chen and colleagues reported significant between-group differences in ARAT, 9HPT (measures fine manual dexterity) and MAL (measures gross arm function).[55] In the ARAT test, the number of patients reaching minimal clinically important difference (MCID) in the experimental group (n=4) was also more than that of the control group (n=1).[55] In addition, 1 study[49] reported significant pre-to-post differences in the experimental groups only, and 3 studies[53,54,57] reported significant within-group differences in both experimental and control groups.

Significance varies within a single study, producing different results by different assessments or subscores. For instance, Chen et al.[54] reported significant between-group differences in the amount of use (AOU) but not in the quality of movement (QOM). This may indicate that the TBS treatment is more effective in improving certain aspects of functions than other aspects.

In conclusion, in the 9 studies reviewed in this part, 5 studies[51,52,55,56,58] reported significant between-group differences, 1 study[49] reported significant outcomes in the experimental group only, and 3 studies[53,54,57] reported significant outcomes in both experimental and control groups. These results suggest that TBS has an overall positive effect on improving function, but since a portion of these studies reported a lack of significance, more evidence is needed to determine whether the effect is stable or strong. The most commonly used assessment in these 9 studies is ARAT. This may suggest that ARAT detects functional effects better than other measures and could be considered in future studies.

2.2.1.2. Sustained Effect of TBS on Function

Four studies[49,53,57,58] in this section designed follow-ups, and 3 studies[53,57,58] reported significant between-group differences or within-group differences at follow-ups. Ackerley et al. reported significant between-group differences in ARAT favoring the experimental group at 1-month post-intervention.[58] In addition, Hsu et al. reported significant improvements in ARAT in all groups from baseline to 60 days post-stroke (4-6 weeks post-intervention),[57] and Talelli et al. reported significant within-group difference in both cTBS and cSham groups at 1-month post-intervention (statistical tests were not conducted at 3 months because of inadequate numbers in the cTBS group), and in both iTBS and iSham groups at 3 months post-intervention.[53]

To sum up, 1 study[58] reported significant between-group differences favoring the experimental group at 1-month post-intervention, and 2 studies[53,57] reported significant within-group improvements in both experimental and control groups up to 3 months after the intervention. This result suggests that the effect of TBS on function may be sustained after the treatment is over, but this evidence is limited so more studies are needed to confirm this conclusion. As commonly designed in present studies, follow-ups at 1 to 3 months after intervention could be considered in future studies.

2.2.2. Effect of TBS on Motor Function

As in research on the effects of traditional rTMS on post-stroke upper extremity recovery, many studies on the effectiveness of TBS measured motor function. Seven[49,51,52,55–58] studies of the 9 RCTs reviewed used motor function outcome measures. The assessments include the Fugl-Meyer Assessment (FMA) and the National Institutes of Health Stroke Scale (NIHSS).

2.2.2.1. Immediate Effect of TBS on Motor Function

Six studies[49,51,52,55–57] reported significant between-group or within-group differences immediately after the TBS intervention. Five studies[49,51,52,56,57] reported significant between-group differences favoring the experimental groups. In 2 of these 5 studies, the researchers coupled iTBS with 1-Hz rTMS and designed 3 experimental groups (the iTBS + rTMS group; the iTBS-only group; the rTMS-only group).[52,56] Sung et al. also added a sham group.[52] Both studies reported better improvements in the combined group compared to iTBS-only and rTMS-only groups. Sung and colleagues reported a significant difference between the combined group and the sham group favoring the combined group as well.[52] Kuzu et al. also included both TBS and rTMS in the study, with 2 experimental groups (1-Hz rTMS and cTBS treatment respectively) and a control. Significant between-group differences were reported both between the cTBS group and the control and between the rTMS group and the control after treatment.[49] While these 3 trials provide evidence for the effects of TBS, they also enhance the evidence for the effects of rTMS that was discussed earlier in this review. In addition, 1 study[55] reported significant within-group improvements in both experimental and control groups, and 1 study[58] reported no significance. Divergence also occurs within a single study. Hsu et al.[57] reported significant between-group differences in NIHSS, while not in FMA.

In summary, in the 7 studies reviewed in this part, 5 studies[49,51,52,56,57] reported significant between-group differences, 1 study[55] reported significant

pre-to-post differences in both the experimental and control groups, and 1 study[58] reported no significance. These results suggest that TBS has potentially positive effects on upper extremities motor function, but the evidence is not strong. As 2 studies[51,57] reviewed are pilot studies that had small sample sizes, it is suggested that future studies could recruit more patients. The FMA is most commonly used in these TBS studies as well, so it could be considered by future studies first when choosing assessments.

2.2.2.2. Sustained Effect of TBS on Motor Function

Three studies[49,57,58] designed follow-ups and two[49,57] reported significant between-group differences favoring the experimental group. Kuzu et al. reported significant between-group differences in FMA between the control and both experimental groups (1-Hz rTMS group and cTBS group) favoring 2 experimental groups at 1-month post-intervention.[49] This result also supports the sustained effect of rTMS on motor function. Hsu and colleagues reported significant between-group differences favoring the experimental group in both FMA and NIHSS at 60 days post-stroke (4-6 weeks post-intervention).[57] The study by Ackerley et al. reported no significance at follow-ups.[58]

In conclusion, 2 studies[49,57] reported significant between-group differences favoring experimental groups at follow-ups and 1 study[58] reported a lack of significance. This result suggests that the effect of TBS on motor function could be sustained for about 1 month after the intervention. Future studies could design follow-ups later than 1 month to examine whether the effect of TBS on motor function could be sustained.

2.2.3. Effect of TBS on Spasticity

Studies on TBS assess the spasticity more frequently than those studying rTMS. Four[49,51,54,55] out of 9 studies used assessments that measure spasticity. The outcome measures include the Modified Ashworth scale (MAS), Modified Tardieu Scale (MTS), and Shear Wave Ultrasound Elastography (SWV).

2.2.3.1. Immediate Effect of TBS on Spasticity

All 4 studies produced statistically significant differences between experimental and control groups. Kuzu et al.[49] designed 2 experimental groups using 1-Hz rTMS and cTBS, respectively. While significant pre-to-post improvements were reported in both experimental groups, between-group difference was reported only between the cTBS and control groups, which favors the cTBS group. Specifically, this significance is on the wrist flexor.[49] This study may suggest that TBS is more effective in improving spasticity in stroke patients than rTMS. Moreover, Chen et al.[54] used 3 different assessments of spasticity and reported significant between-group differences favoring the experimental group in all 3 assessments. These assessments measured spasticity of the elbow, wrist, bicep brachii, and carpi radialis. This result suggests that improvements in spasticity after TBS may exist across multiple arm muscles.[54]

In conclusion, all 4 studies[49,51,54,55] reviewed reported significantly better outcomes in the experimental group. This result suggests that TBS indeed has a positive effect on spasticity. Also, more studies of TBS examining this aspect and the more positive results it produced compared to the rTMS that was examined above may suggest that TBS is more effective in improving spasticity

than rTMS.

2.2.3.2. Sustained Effect of TBS on Spasticity

There was only 1 study[49] on spasticity that designed follow-ups and this study reported significant between-group differences favoring the experimental groups. Kuzu and colleagues designed 2 experimental groups using cTBS and rTMS, respectively. Both experimental groups produced significantly better outcomes 4 weeks after intervention in MAS wrist flexor scores when compared to the control.[49] This result also strengthens the evidence for the sustained effect of rTMS on spasticity. Since only 1 study reported evidence of the sustained effect of TBS, future studies could conduct more follow-ups to fill this blank.

2.2.4. Effect of TBS on Muscle Strength

Two studies[52,53] in this section examined the effect of TBS on muscle strength. The assessments include the Medical Research Council (MRC) and hand grip strength (HGS). One study[52] reported statistically significant differences between groups. Sung et al. used finger flexor MRC and designed 3 experimental groups (1-Hz rTMS + iTBS; iTBS only; rTMS only) and a control group. All 3 experimental groups produced significantly better outcomes when compared to the control group. Additionally, significant between-group differences were reported between the group combining 1-Hz rTMS and iTBS and each of the iTBS-only and rTMS-only groups, both favoring the combined group.[52] This result supports both the effect of TBS and of rTMS on improving muscle strength. Talelli et al.[53] tested grip strength and reported no significance except for an effect of time on the grasp strength. No significant results were reported at follow-ups. To sum up, these 2 studies produced mixed results and the number of studies reviewed is limited, so the result is inconclusive on the effect that TBS has on muscle strength. More evidence is needed to provide an accurate perspective. Future studies may choose to use MRC or HGS, and they should explore which one is better and use that assessment more often.

2.2.5. Effect of TBS on Quality of Life

In studies on the effect of TBS, 1 study[55] reviewed examined patients' improvement in QOL using the Stroke Impact Scale (SIS). In this RCT, a statistically significant difference is reported between the experimental and control groups favoring the experimental group. Moreover, 2 patients reached MCID in the experimental group, while none did in the control group. No follow-up tests were designed in this study. This result supports the immediate effect of using TBS to improve QOL in stroke rehabilitation. However, the evidence is limited and the sample size is relatively small (n=23) in this study.[55] As a result, future studies can further investigate the effect of TBS on QOL, since this study shows potential effects. Also, studies with larger sample sizes could provide more credible evidence for the effects of TBS on QOL.

2.2.6. Effect of TBS on Cortical Excitability

In the 9 studies reviewed in the TBS section, 5 studies[52,54,56–58] measured cortical excitability using the measurement of motor evoked potential (MEP) amplitude and latency, laterality index (LI), and motor map area.

2.2.6.1. Immediate Effect of TBS on Cortical Excitability

Three studies[52,54,56] reported significant between-group differences or within-group differences in this section. Meng et al.[56] and Sung et al. both designed a combination of 1-Hz rTMS and iTBS with 3 experimental groups (1-Hz rTMS + iTBS; rTMS only; iTBS only). Sung et al. also designed a control. Sung and colleagues reported statistically significant differences in the contralesional motor map area favoring the experimental group, between the control and three experimental groups, respectively. Significant differences between the combined and control groups in MEP amplitude and latency were also reported.[52] Meng et al. reported significantly better results in MEP amplitude and latency between the combined group and the rTMS-only group in the extensor digitorum communis (EDC) and biceps brachii, and between the combined group and the iTBS-only group in abductor brevis pollicis (ABP), EDC and biceps brachii. Moreover, 1 study[54] reported significant improvement of MEP in the experimental group, while only 2 studies[57,58] reported no significance. No significance is reported at follow-ups.

In summary, in the 5 studies[52,54,56–58] reviewed in this section, 2 studies[52,56] reported significant between-group differences, 1 study[54] reported significant within-group differences in the experimental group only, and 2 studies[57,58] reported no significance immediately after the intervention. The results of cortical excitability after rTMS are mixed. There is a potential effect that TBS may help regain the interhemispheric balance, but the lack of significance in 2 studies means further evidence is needed to confirm the effect of TBS in modulating cortical excitability. As all 4 studies reviewed used MEP to measure cortical excitability, future studies could also consider using this measurement.

2.3. Comparison of Facilitatory and Inhibitory rTMS

Both rTMS and TBS have two subtypes, facilitatory (HF-rTMS and iTBS) and inhibitory (LF-rTMS and cTBS). In this section, the usage of facilitatory rTMS and inhibitory rTMS will be discussed and the efficacy of these two types of rTMS on function and motor function compared.

There are differences between the number of studies of facilitatory rTMS and inhibitory rTMS. In the 20 RCTs reviewed, 14 studies[22,23,25–27,34,51–58] used facilitatory rTMS and 12 studies[21,22,24,25,28–30,34,49,52,53,56] used inhibitory rTMS. The overall evidence shows a balanced number of studies on facilitatory and inhibitory types of rTMS. However, 11 studies[21,22,24,25,28–30,34,49,52,56] used LF-rTMS while only 6 studies[22,23,25–27,34] used HF-rTMS; 8 studies[51–58] used iTBS while only 2 studies used cTBS. In summary, while facilitatory rTMS and inhibitory rTMS are equally commonly used, LF-rTMS is used more often than HF-rTMS and iTBS is used more than cTBS.

This difference in a number of studies may be due to the perceived differences in effectiveness or adverse effects of each application. Xiang et al. reported better positive effects on motor recovery in the LF-rTMS group, although no significant differences between different stimulation frequency groups.[59] The systematic review by Zhang et al. noted that iTBS has a greater significant mean effect size than cTBS (0.60 and 0.35 respectively).[18] A recent meta-analysis concluded that LF-rTMS may be more beneficial than HF-

rTMS.[60] Another reason why there are more LF-rTMS studies may be safety. Wassermann et al. reported seizure risks in patients who received HF-rTMS.[61] This was hypothesized to be due to the higher frequency and intensity. The safety and efficacy of the technique are likely to contribute to the number of studies on the technique.

In the trials reviewed in this study, there were generally no better effects on function or motor function in HF-rTMS versus LF-rTMS. Four[22,23,25,26] of 5 studies of functional change following HF-rTMS reported significant between-group differences favoring HF-rTMS groups and one[27] reported significant pre-to-post differences in the experimental group only. Seven[22,24,25,28,30,52,56] of 10 studies of functional changes following LF-rTMS reported significant between-group differences favoring the experimental group, one[29] reported significant pre-to-post differences in the experimental group only, and two[21,49] reported significant pre-to-post differences in both experimental and control groups. These results suggest that the efficacy of LF-rTMS and HF-rTMS on function is equivocal. Four[22,23,26,34] of 5 studies of motor function following HF-rTMS reported significant between-group differences favoring the experimental group and one[27] reported significant pre-to-post differences in the experimental group only. Seven[22,24,28,34,49,52,56] of 10 studies of motor function following LF-rTMS reported significant between-group differences favoring the experimental group, one[29] reported significant pre-to-post differences in the experimental group only, one[21] reported significant pre-to-post differences in both groups, and one[30] reported a lack of significance. This suggests no tendency advantage for LF-rTMS in motor function. In summary, LF-rTMS and HF-rTMS have been reported to be equally effective on function and motor function.

The facilitatory form of TBS, iTBS, is reported to be more effective than cTBS in function. Five[51,52,55,56,58] of the 8 studies of functional change following iTBS reported significant between-group differences favoring the experimental group and 3[53,54,57] reported significant pre-to-post differences in both experimental and control groups. One[49] of the 2 studies of functional change following cTBS reported significant pre-to-post differences in the experimental group only, and the other one[53] reported significant pre-to-post differences in both experimental and control groups. This result does suggest better efficacy in functional recovery in the iTBS studies. Four[51,52,56,57] of 6 studies of motor function following iTBS reported significant between-group differences favoring the experimental group, one[55] reported significant pre-to-post differences in both experimental and control groups, and one[58] reported a lack of significance. Only one study[49] assessed motor function following cTBS and reported a significant between-group difference favoring the experimental group. Therefore, the evidence is insufficient for the effect of cTBS on motor function, and it is inconclusive whether iTBS is more beneficial to motor function than cTBS. In conclusion, iTBS is more effective in improving function than cTBS while no conclusion can be drawn about the effect of iTBS and cTBS on motor function.

In summary, roughly the same number of studies examined facilitatory and inhibitory rTMS. However, LF-rTMS is studied more frequently than HF-rTMS, and the iTBS is studied more than cTBS. One reason may be better efficacy, as concluded by several systematic reviews and meta-analyses[18,59,60] and by this review. Facilitatory TBS (iTBS) has been reported to be more

effective than inhibitory TBS (cTBS) on function, while no other trend favoring either facilitatory rTMS or inhibitory rTMS is reported. Future studies should replicate findings in function and motor function following HF-rTMS, as consistent results were reported in previous studies. For cTBS, the results of the 2 studies reviewed were not consistent in functional changes; this may be due to different outcome measures used or unmatched dosage. As a result, future studies should try to use a unified dosage as well as comparable outcome measures. Studies examining function and motor function outcomes following HF-rTMS and cTBS and comparing facilitatory and inhibitory rTMS will help inform clinical decision-making. Providers seeking to improve the patients' functional abilities may wish to choose iTBS versus cTBS because of the superior functional outcomes noted in this review.

2.4. Adverse Effects

In the studies that examined the effects of post-stroke interventions, adverse effects are often reported and may indicate potential safety concerns in clinical practice. In the following two subsections, the adverse effects reported in RCTs that used rTMS and TBS are summarized and discussed.

2.4.1. Adverse Effects of rTMS

In the 11 studies, 4 studies[21,24,28,34] reported adverse events. A total of 140 adverse events were reported in the 696 patients. These adverse events are mostly mild effects that include transient headaches, arm and hand pain, and so on. Harvey et al. noted in their study (n=196) that all 26 events in the study were resolved within 24 hours.[21] Kim et al.[24] reported the greatest number of adverse events, a total of 111 events in 77 patients. 105 of them are mild events and 7 events were due to the treatment device. There were 71 events in the 2 experimental groups and 40 events in the control group, and the study found no significant difference in adverse effects between the experimental and the control group. However, 2 serious adverse events happened in each of 2 studies. Sharma et al. reported 1 event of seizure developed 18 hours after the fourth treatment session[28]; Kim et al. reported 1 ischemic stroke reoccurrence, but it was determined to be unrelated to the intervention.[24]

To sum up, there is a certain proportion of adverse events in RCTs reviewed that use rTMS in stroke rehabilitation. However, since most events are transient or slight, the side effects of rTMS could be considered generally mild, which is also consistent with the non-invasive feature that makes it favorable for stroke rehabilitation. Still, as there was 1 event of seizure in the studies reviewed, the rTMS intervention should be administered carefully to appropriate patients.

2.4.2. Adverse Effects of TBS

In 9 studies reviewed in this section, 3 studies[55–57] reported adverse events. A total of 8 adverse events were reported in the 250 patients. All of the adverse events are mild events that include light nausea, mild headache, transient local pain, tingling sensations, and muscle soreness. In addition, the case of muscle soreness was reported by Chen et al. after the virtual reality-based cycling training (VCT) in the study.[55] Therefore, the adverse event may not be directly related to the TBS treatment that the patient received, but is more likely to be

the result of the VCT treatment that was combined with the TBS intervention. To sum up, the adverse events of TBS do tend to be less frequent and milder than that of rTMS. This result is consistent with the assumption of some researchers that TBS may be a more comfortable alternative for conventional rTMS.[49] However, it is also observed that the studies reviewed in the TBS section mostly have smaller sample sizes than studies in the rTMS section. Therefore, more studies in TBS using larger sample sizes should be done to reconfirm this assumption.

2.5. Appropriate/Inappropriate Patients for rTMS

This section will discuss appropriate and inappropriate patients for rTMS based on inclusion and exclusion criteria as well as the properties of rTMS.

2.5.1. Appropriate Patients

In the 20 trials reviewed using rTMS and TBS, there are common inclusion criteria, which could be indicative of the standard for suitable patients. Eighteen studies[21–23,25–30,34,49,51–54,56–58] require patients to be within a specific time range post-stroke; 14 studies[21,24–30,49,51,54–57] set age ranges for patients included; 12 studies[22–24,26,28,34,51,53–55,57,58] included only patients with no previous stroke history; 7 studies[24,28,30,34,49,53,57] stated that only patients with certain type of stroke are eligible for inclusion; 6 studies[22,23,34,55,57,58] listed specific stroke lesion location as inclusion criteria. Sixteen studies[21,24,26–30,34,49,52–58] also set standards of assessment score ranges for eligible candidates. However, these criteria are all set only for research purposes and do not need to be considered when applied to clinical therapies. RCTs include these criteria to make sure that the patients' conditions are similar and to reduce the influence of other variables, and therefore make the results more comparable, while in the clinical setting, these conditions do not affect the application of a treatment. In addition to these research-only criteria, 4 studies[26,29,53,54] stated that only patients with intact cognitive abilities were considered appropriate candidates and 10 studies[22,23,25,27,30,34,51,52,55,56] mentioned patients with cognitive impairments in their exclusion criteria. This criterion is important in both research settings and clinical practices. Since the application of rTMS requires patients to follow the instructions of the therapist, the ability of patients to clearly understand the orders of the therapist is indispensable.

Other factors regarding appropriate patients of rTMS include the patient's accessibility to a hospital and funds available. Applying rTMS requires technical equipment and the assistance of trained medical technicians. Therefore, it is not a treatment method that can be carried out by the patients themselves at home. As a result, patients who have the mobility to get access to hospitals or who have caregivers who can assist them to go to the hospital are more suitable for rTMS. Additionally, as the application of rTMS is expensive (207.24 USD per session),[62] patients with health funds or medical insurance are more appropriate candidates.

In summary, the rTMS treatment, as non-invasive brain stimulation, unlike the physical therapies, does not require patients to possess a great level of motor function or muscle strength. However, appropriate patients should indeed have unimpaired cognitive abilities. Furthermore, patients who can

afford the treatment costs and have access to clinical institutes and therapists are appropriate for rTMS.

2.5.2. Inappropriate Patients

Similar to inclusion criteria, there are also commonalities in exclusion criteria of these 20 trials using rTMS and TBS. Seventeen studies[21–30,34,49,51,53,55,56,58] excluded patients with TMS or MRI contraindications, including pacemakers, intracranial metallic devices and cochlear implants. This criterion is of great significance to ensure safety because rTMS uses magnetic fields and generates electric currents in the patient's brain, and metallic implants like pacemakers could be susceptible to the magnetic field. Therefore, patients with such conditions are unsuitable for rTMS treatment. Thirteen studies[21,24,26–29,49,51–53,55–57] stated that patients with epilepsy and seizure histories would not be eligible. This is also an important standard to classify a patient as inappropriate for rTMS. As seizures are caused by uncontrolled electrical disturbances in the brain, the electric currents that rTMS generates during the intervention could have the risk of causing the patient to experience seizures. Such an adverse event was also reported in one study.[28] Moreover, 13 studies[21,22,29,34,49,51–58] excluded patients with other neurological diseases. This criterion was used in studies to avoid introducing variables that may affect the interpretation of outcomes; however, in a clinical situation, patients with some neurological conditions that are directly associated with the brain may not be suitable for rTMS, since such conditions may complicate the situation and cause potential side effects, while patients with other conditions may not be unsuitable for rTMS. As a result, whether the patient is inappropriate for rTMS depends on the exact neurological condition he or she has, and in such cases, the physician should determine whether the patient's condition is contradicted with rTMS intervention. In addition, 12 studies[21,23,24,27,28,30,49,52,54–57] excluded patients with other medical conditions such as pregnancy and history of other diseases and 8 studies[25,26,30,49,54–57] excluded patients with serious complications such as functional decline or failure of organs. These two criteria are meaningful in both research settings and clinical practices, as these factors make patients' conditions more complex and may affect both the safety of applying rTMS and the interpretation of results for research purposes. Moreover, in such cases, maintaining a stable health condition is the primary concern of patients, while treating arm dysfunctions caused by stroke onset is often not their top priority. Lastly, 10 studies[22,26,27,29,34,52,55–58] clarified that patients with aphasia were not eligible. This criterion is useful in studies because researchers would want to exclude patients who have potential communicative problems; however, in clinical settings, clinicians should determine if the patient is able to understand and follow instructions and finish the treatment.

In conclusion, patients who have TMS contraindications such as metallic implants, histories of epilepsy, concomitant neurological disorders, other inconvenient conditions, or poor general health are unsuitable for rTMS treatment.

2.6. Combination of rTMS with Other Treatments

rTMS is frequently coupled with other treatments and multiple comparisons conducted have reported its effects in strengthening the results of other interventions. Chen et al. combined iTBS and virtual reality-based cycling training (VCT) and concluded that iTBS may be a safe option that has great potential in enhancing the effects of conventional rehabilitation as an adjuvant therapy.[55] Miu et al. also stated that rTMS may augment the effects of traditional neurorehabilitation.[63]

In the studies reviewed, 16 studies combined rTMS and other therapies in their study protocols. Thirteen studies[21,22,24–26,28–30,49,51,53–58] combined rTMS with physical therapy or conventional rehabilitation, which are well-established and widely applied therapies in stroke rehabilitation. This also reflects the fact that therapies and interventions are often combined in stroke rehabilitation to produce better and more comprehensive effects.[64] Moreover, Kim et al. combined rTMS with task-based occupational therapy, Pinto et al. combined rTMS with fluoxetine, and Chen et al. combined TBS with VCT.[24,30,55] Significant between-group differences favoring experimental groups in functional outcomes were reported in all 3 studies;[24,30,55] Chen et al. also reported significantly better results in the experimental group in motor function, spasticity, and quality of life,[55] and Kim et al. reported significantly better improvements in the experimental group in motor function and muscle strength.[24] These results suggest that combinations of rTMS with other interventions that focus on different aspects after stroke may be prospective options in stroke rehabilitation. Future studies could couple rTMS with other interventions that are not yet combined to investigate potential augmented effects, or do further research on the combinations that have already been used to verify their effectiveness.

3. Comparison of rTMS with Transcranial Direct Current Stimulation (tDCS)

3.1. Overview of tDCS

Transcranial Direct Current Stimulation (tDCS) is another form of non-invasive brain stimulation. In tDCS, mild electrical currents (1-2mA) are conducted to the brain through two electrodes placed over the area of interest in the brain. [65] The tDCS intervention modulates cortical excitability.[65] There are two types of tDCS—anodal tDCS and cathodal tDCS. Anodal tDCS increases cortical excitability and is targeted to the lesioned hemisphere. Cathodal tDCS targets the unaffected hemisphere to decrease cortical excitability.[65] Table 4 summarizes the information about studies evaluating the effects of tDCS on the upper limb post-stroke.

Table 4. *Summary of studies of tDCS*

Author, year study design sample size time post stroke category	Interventions Duration: session length, frequency per week for total number of weeks	Outcome measures Results (direction of effect)
Alisar et al., 2020 **RCT** **N_{start}=32** **N_{end}=32** **TPS=subacute & chronic**	E: bihemispheric tDCS C: sham tDCS Duration: 30min/d, 5d/wk, 5wks + PT & OT (30-90min)	• FMA: E (+post, ***), C (-); E vs C (-) • FIM: E (+post, ***), C (-); E vs C (+exp, *) • BRS: E (+post, ***), C (-); E vs C (-)
Bolognini et al., 2020 **RCT** **N_{start}=32** **N_{end}=21** **TPS=acute**	E: bihemispheric tDCS C: sham tDCS Duration: 2 sessions/d, 5 days in total	• HGS: E & C (+post, ***) 2 months and 6 months; E vs C (-) • MI: E & C (+post, ***) at EOT, 2 months and 6 months; E vs C (-) • NIHSS: E & C (+post, ***) at EOT, 2 months and 6 months; E vs C (-) • BI: E & C (+post, ***) at EOT, 2 months and 6 months; E vs C (-)
Bornheim et al., 2020 **RCT** **N_{start}=50** **N_{end}=46** **TPS=acute**	E: anodal tDCS C: sham tDCS Duration: 20min/d, 5d/wk, 4wks + 2h conventional rehabilitation	• WMFT: E & C (+post, ***) at EOT, 3 months, 6 months and 1 year; E vs C (+exp, *) at EOT, 3 months, 6 months and 1 year • FMA: E & C (+post, ***) at EOT, 3 months, 6 months and 1 year; E vs C (+exp, *) at 6 months and 1 year • SWMT: E & C (+post, ***), E vs C (+exp, *) at EOT, 3 months, 6 months and 1 year • BI: E & C (+post, ***) at EOT, 3 months, 6 months and 1 year; E vs C (-) • SIS: E & C (+post, ***) at EOT, 3 months, 6 months and 1 year; E vs C (-)

Author, year study design sample size time post stroke category	Interventions Duration: session length, frequency per week for total number of weeks	Outcome measures Results (direction of effect)
Lefebvre et al., 2014 **RCT** **N_{start}=19** **N_{end}=19** **TPS=chronic**	E: bihemispheric tDCS C: sham tDCS Duration: 20min/session, 2 sessions in total, at least one week apart	• PLD: E (+post, *), C (-); E vs C (-) • ULD: E (+post, ***), C (-); E vs C (-) • GFI/LFI: E & C (+post, *); E vs C (-) • PPT: E (+post, ***), C (-) ; E vs C (+exp, ***)
Kim et al., 2021 **RCT** **N_{start}=30** **N_{end}=30** **TPS=chronic**	E: bihemispheric tDCS C: sham tDCS Duration: 20min/d, 5d/wk, 4wks + mCIMT	• FMA: E & C (+post, ***); E vs C (-) • MAL: AOU: E (+post, ***), C (-); E vs C (+exp, *) QOM: (-) use of unaffected side: E (+post, ***), C (+post, **); E vs C (-) use of affected side: E & C (+post, ***); E vs C (-)
Rocha et al., 2016 **RCT** **N_{start}=21** **N_{end}=21** **TPS=chronic**	E1: cathodal tDCS E2: anodal tDCS C: sham tDCS Duration: 9 or 13min/sessions, 3d/wk, 4wks + 6h/d mCIMT	• FMA: E1, E2 (+post, *) at EOT and at 1 month, C (-) ; E2 vs C (+exp, *), E1 vs C (-), E1 vs E2 (-) • MAL: AOU: E1, E2 (+post, *) at EOT; E1, E2, C (post, *) at 1 month; between groups: (-) QOM: E1, E2 (+post, *) at EOT and 1 month, C (-) • HGS: E(-), C (+post, *) at 1 month

Author, year study design sample size time post stroke category	Interventions Duration: session length, frequency per week for total number of weeks	Outcome measures Results (direction of effect)
Lee and Chun, 2014 **RCT** **N_{start}=59** **N_{end}=54** **TPS=subacute**	E1: cathodal tDCS E2: VR E3: cathodal tDCS + VR Duration: 30min/d, 5d/wk, 3wks + conventional rehabilitation	• MMT: shoulder: E1, E2, E3 (+post, *) wrist: E1 (+post, *), E2 & E3 (-) between group: (-) • MFT: E1, E2, E3 (+post, *); E3 vs E1 (+exp3, *), E3 vs E2 (+exp3, **), E1 vs E2 (+exp2, **) • FMA: E1, E2, E3 (+post, *); E3 vs E1 (+exp3, *), E3 vs E2 (+exp3, **), E1 vs E2 (+exp2, *) • MAS: (-) • BBT: E3(+post, *), E1 & E2 (-); between group: (-) • K-MBI: E1, E2, E3 (+post, *); between group: (-)
Yao et al., 2020 **RCT** **N_{start}=42** **N_{end}=40** **TPS=subacute**	E: cathodal tDCS C: sham tDCS Duration: 20min/d, 5d/wk, 2wks, together with 20min VR therapy	• FMA: E & C (+post, ***); E vs C (+exp, **) • ARAT: E & C (+post, ***); E vs C (+exp, *) • BI: E & C (+post, ***); E vs C (+exp, *)

Abbreviations and table notes: C=control group; E=experimental group; h=hours; min=minutes; d=days; wks=weeks; RCT=randomized controlled trial; TPS=time post stroke category (acute: less than 1 month, subacute: more than 1 month but less than 6 months, chronic: more than 6 months); EOT=end of treatment; M1=primary motor cortex; UL=upper limb; VT: virtual reality; OT=occupational therapy; MI=Motricity Index; AOU=amount of use; QOM=quality of movement; mCIMT=modified constraint-induced movement therapy;
"+post" indicates a significant within-group difference in favor of post-treatment scores
"+pre" indicates a significant within-group difference in favor of baseline scores
"+exp" indicates a significant between-group difference in favor of the experimental group
"+exp2" indicates a significant between-group difference in favor of the first experimental group
"+exp3" indicates a significant between-group difference in favor of the third experimental group
"+con" indicates a significant between-group difference in favor of the control group
"-" indicates no significant within-group/between-group difference
"*" indicates a significance level of $p<0.05$; "**" indicates a significance level of $p<0.01$; "***" indicates a significance level of $p<0.001$.
"at xx months/wks" means that the result is tested at xx months or weeks post-intervention

3.2. Comparison of rTMS with tDCS

This section will compare rTMS and tDCS in aspects of estimated costs, equipment, duration, assistance needed, mechanism of effect, appropriate/inappropriate patients, and efficacy in stroke rehabilitation of the upper extremities. In addition, the similarities and differences between the two interventions will be summarized and discussed.

3.2.1. Similarities between rTMS and tDCS

There are some similarities between rTMS and tDCS. Both rTMS and tDCS are non-invasive brain stimulation methods that modulate cortical excitability and produce effects based on the interhemispheric competition model. Both treatments include two subtypes, producing effects of increasing cortical excitability in the affected hemisphere and decreasing cortical excitability in the unaffected hemisphere. Moreover, both treatments require devices and need to be administered by a trained therapist in the hospital. A common advantage of these two treatments is that adverse effects reported are relatively mild. Both treatments require the patient to be able to follow the instructions of the therapist and access a hospital and professional clinicians. As each intervention generates currents through the brain, patients with metal implants or a history of epilepsy are unsuitable for both treatments.

rTMS and tDCS also have similarities in the reported effects on the upper extremities post-stroke. Studies of both treatments reported significant between-group differences favoring experimental groups in function, motor function, and spasticity. In an RCT comparing rTMS and tDCS, equivocal results in activities of daily living (ADL) improvements of the two treatments tested by mBI were also reported.[63]

In summary, both rTMS and tDCS modulate cortical excitability and are performed at hospitals with the assistance of a therapist. The two treatments are also relatively safe and have similar appropriate/inappropriate patients. Moreover, both interventions have effects on function, motor function, and spasticity.

3.2.2. Differences between rTMS and tDCS

There are some aspects in which tDCS differs from rTMS. tDCS is estimated to be less expensive than rTMS; a tDCS session is also shorter (about 20 minutes) than an rTMS session (30~40 minutes). In addition, although both treatments are based on restoring the interhemispheric balance through inhibition or excitation, the mechanisms of modulation of the two treatments are different.[66] While tDCS modulates the neuron membrane potentials and activates a larger number of neurons, rTMS induces a more focal electric field and then generates action potentials in a specific neural circuit.[65,67,68] This makes rTMS more suitable for stimulation of specific white matter tracts.[63]

There are also differences between rTMS and tDCS in their efficacy. The rTMS studies reported significant between-group differences favoring the experimental groups in muscle strength and quality of life, while studies of tDCS only reported significant within-group differences. The RCT comparing rTMS and tDCS reported that rTMS may have better effects on fine hand motor function recovery than tDCS.[63] However, one study of tDCS reported a between-

group difference favoring the experimental group in sensory function, which was used by no studies of rTMS reviewed. This may indicate the potential effects of tDCS on sensory function.

In summary, tDCS is less expensive and has a shorter duration than rTMS, while rTMS has a more focal effect on the brain. Regarding their effects on stroke rehabilitation, rTMS might be more effective in muscle strength and quality of life, and tDCS may have a potential effect on the sensory function which was not reported in rTMS studies.

Table 5. *Comparison of rTMS and tDCS by estimated costs, equipment, duration, where it can be performed and assistance needed*

	Estimated costs	Equipment	Duration	Where it can be performed and assistance needed
Repetitive Transcranial Magnetic Stimulation (rTMS)	207.24 USD per session[62]	an external device that delivers repetitive pulsed magnetic fields of sufficient magnitude[69]	30-40min per session	• at a hospital • professional therapist needed
Transcranial Direct Current Stimulation (tDCS)	167.72 USD per session[62]	a battery-powered current generator connected with two sponge-based electrodes that deliver direct currents to the brain[62]	~20min per session	• at a hospital • professional therapist needed

Abbreviation: USD=United States Dollar

4. Discussion

4.1. Primary Findings

The two forms of rTMS (rTMS and TBS) have been reported to have effects on function, motor function, muscle strength, spasticity, and quality of life post-stroke. Evidence for positive effects of rTMS was especially strong in function, motor function, and spasticity. Sustained effects of rTMS were also reported in function, motor function, muscle strength, and spasticity for up to 1 year. As a brain stimulation method based on the modulation of interhemispheric balance, rTMS was shown to promote beneficial cortical changes, and correlations between cortical excitability and motor function changes were reported.[22,34] Adverse effects of rTMS are generally mild, although there is a risk of seizures after treatment. Studies combining rTMS with other interventions have reported augmented effects of rTMS. The result suggests that rTMS is safe and

overall effective in different aspects of stroke rehabilitation, and it is an ideal adjuvant therapy to be applied in combination with other therapies.

In comparison with tDCS, advantages of each treatment were found. While rTMS may have more focal effects and be more effective in muscle strength and quality of life, tDCS has the benefit of lower costs and a shorter duration, as well as a potential in improving sensory function. Therefore, in clinical practices, physicians could choose treatment according to their goals.

4.2. Unanswered Questions

Although rTMS was found to have potential effects in multiple aspects and criteria in this review, there are still several limitations. Some of the studies reviewed used relatively small sample sizes ($n<20$) or are pilot studies,[51,57,58,70] which makes the results less rigorous. Future studies should select the outcome measures that can detect changes the best and are commonly used, and do power analyses to ensure the sample size is sufficient. In addition, the studies reviewed included patients of acute, subacute, and chronic phases, but the different phases of patients were not discussed in this review. There may be differences among the effects of rTMS on patients of different stages since the state of the brain varies across phases. rTMS may be most effective in the early stages after stroke, as in the acute phase, the cortical reorganization is more active and rTMS exerts effects by influencing the brain. Duration and protocol also vary across studies reviewed, which may introduce variables that could affect the interpretation of results in this review. The optimal parameters are not yet specified as well.

Systematic reviews and meta-analyses could be conducted to help inform these two problems. A future study could divide the RCTs reviewed by stages into three groups—acute, subacute and chronic, conduct statistical tests to evaluate the effect sizes of each group, and then compare to see whether there are significant differences among the three groups. The selection of studies for this review needs to ensure that the stroke type, intervention protocol and outcome measures used in the RCTs are the same. For the protocol problem, the future review could divide the RCTs into several major groups by different parameters and then conduct comparisons and statistical tests. Through this process, a protocol with the most significant effects could be determined, which would be meaningful in clinical applications and help future studies unify their protocols.

Besides limitations, there are other unanswered questions. Pain, abnormal synergy, and sensory functions were not measured in the studies reviewed. Therefore, the effect of rTMS in these aspects is still unclear. The lack of research may be because of the absence of appropriate outcome measures. Sensory function and pain are both perceptions by the patient and are difficult to measure objectively; abnormal synergy is also difficult to assess for its complexity, while currently, a subscore of NIHSS tests it. It is reasonable to assume that rTMS may be effective in abnormal synergy, as coordination of muscles is related to brain activities. However, fewer effects could be assumed about the effects of rTMS on reducing pain, as pain induction is more biomechanical; as a result, future studies could possibly apply a package of treatments that target both the brain and the joint itself. Future studies on

designing self-rating scales for pain and sensory function could be helpful to know how patients feel. Moreover, agreement on outcome measures is important so that studies are comparable to each other, so future studies could also focus on determining the best assessment within these aspects.

Lastly, the underlying mechanisms of the effects of rTMS are not completely understood and are yet to be elucidated.[18,22,29,57,71] While the interhemispheric model is widely accepted as the principles of rTMS, there is a competing vicariation model that believes inhibiting the unaffected hemisphere would be counterproductive as compensatory activities of the contralesional side are interfered with.[72] Di Pino et al. proposed a new model for brain plasticity which combines these two models by introducing a new parameter, structural reserve, which describes the degree to which improvements in neural pathways and relays contribute to the recovery in a patient.[72] If the structural reserve is high, the patient's improvement could be better predicted by the interhemispheric competition model; on the other hand, if the structural reserve is low, the vicariation model would better describe the patient's conditions.[72] Based on this model, a hypothesis could be made that some studies reported a lack of significance that may be due to the lower structural reserve of some patients and the differences among patients in structural reserve. This might also suggest that some patients would have better effects following rTMS than other patients, which could be considered in the future by physicians choosing interventions for their patients. Future studies could develop methods to quantify the structural reserve and conduct RCTs to warrant this model and hypothesis to shed more light on the mechanism through which rTMS produces effects.

5. Conclusion

In this review, rTMS was shown to be effective in improving function, motor function, muscle strength, spasticity, and quality of life for rehabilitation of upper extremities in stroke patients. Sustained improvements and correlations between cortical changes induced by rTMS and better motor recovery have also been reported. With relatively mild adverse effects, rTMS could be an ideal choice for stroke rehabilitation or adjuvant therapy. tDCS has similar effects to rTMS while having the advantage of lower costs and shorter treatments; however, treatment choices should be made based on the specific clinical goals for each patient.

References

1. Donkor ES. Stroke in the 21st Century: A Snapshot of the Burden, Epidemiology, and Quality of Life. *Stroke Research and Treatment*. 2018;2018:1-10. doi:10. 1155/2018/3238165
2. Tsao CW, Aday AW, Almarzooq ZI, et al. Heart Disease and Stroke Statistics—2022 Update: A Report From the American Heart Association. *Circulation*. 2022;145(8). doi:10.1161/CIR.0000000000001052
3. Kwakkel G, Kollen BJ, van der Grond J, Prevo AJH. Probability of Regaining Dexterity in the Flaccid Upper Limb: Impact of Severity of Paresis and Time

Since Onset in Acute Stroke. *Stroke*. 2003;34(9):2181-2186. doi:10.1161/01.STR.0000087172.16305.CD
4. Doan QV, Brashear A, Gillard PJ, et al. Relationship Between Disability and Health-Related Quality of Life and Caregiver Burden in Patients with Upper Limb Poststroke Spasticity. *PM&R*. 2012;4(1):4-10. doi:10.1016/j.pmrj.2011.10. 001
5. Mayo NE, Wood-Dauphinee S, Côté R, Durcan L, Carlton J. Activity, participation, and quality of life 6 months poststroke. *Archives of Physical Medicine and Rehabilitation*. 2002;83(8):1035-1042. doi:10.1053/apmr.2002.33984
6. Hirakawa Y, Takeda K, Tanabe S, et al. Effect of intensive motor training with repetitive transcranial magnetic stimulation on upper limb motor function in chronic post-stroke patients with severe upper limb motor impairment. *Topics in Stroke Rehabilitation*. Published online May 2, 2018:1-5. doi:10.1080/ 10749357.2018.1466971
7. Larsen DS, ed. *Neurologic Rehabilitation: Neuroscience and Neuroplasticity in Physical Therapy Practice*. McGraw-Hill Education; 2016.
8. Fisicaro F, Lanza G, Grasso AA, et al. Repetitive transcranial magnetic stimulation in stroke rehabilitation: review of the current evidence and pitfalls. *Ther Adv Neurol Disord*. 2019;12:175628641987831. doi:10.1177/1756286419878317
9. Hummel FC, Cohen LG. Non-invasive brain stimulation: a new strategy to improve neurorehabilitation after stroke? *The Lancet Neurology*. 2006;5(8):708-712. doi:10.1016/S1474-4422(06)70525-7
10. Volz LJ, Grefkes C. Basic Principles of rTMS in Motor Recovery After Stroke. In: Platz T, ed. *Therapeutic RTMS in Neurology*. Springer International Publishing; 2016:23-37. doi:10.1007/978-3-319-25721-1_3
11. Buonomano DV, Merzenich MM. Cortical Plasticity: From Synapses to Maps. *Annu Rev Neurosci*. 1998;21(1):149-186. doi:10.1146/annurev. neuro.21.1.149
12. Abo M, Kakuda W. *Rehabilitation with RTMS*. Springer International Publishing; 2015. doi:10.1007/978-3-319-20982-1
13. Cooke SF. Plasticity in the human central nervous system. *Brain*. 2006;129(7):1659-1673. doi:10.1093/brain/awl082
14. Huang YZ, Edwards MJ, Rounis E, Bhatia KP, Rothwell JC. Theta Burst Stimulation of the Human Motor Cortex. *Neuron*. 2005;45(2):201-206. doi:10.1016/j.neuron.2004.12.033
15. Schwippel T, Schroeder PA, Fallgatter AJ, Plewnia C. Clinical review: The therapeutic use of theta-burst stimulation in mental disorders and tinnitus. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*. 2019;92: 285-300.doi:10.1016/j.pnpbp.2019.01.014
16. Huang YZ, Chen RS, Rothwell JC, Wen HY. The after-effect of human theta burst stimulation is NMDA receptor dependent. *Clinical Neurophysiology*. 2007;118(5):1028-1032. doi:10.1016/j.clinph.2007.01.021
17. Herweg NA, Solomon EA, Kahana MJ. Theta Oscillations in Human Memory. *Trends in Cognitive Sciences*. 2020;24(3):208-227. doi:10.1016/j.tics.2019.12.006
18. Zhang L, Xing G, Fan Y, Guo Z, Chen H, Mu Q. Short- and Long-term Effects of Repetitive Transcranial Magnetic Stimulation on Upper Limb Motor

Function after Stroke: a Systematic Review and Meta-Analysis. *Clin Rehabil*. 2017;31(9):1137-1153. doi:10.1177/0269215517692386
19. Grosset D. Clinical, radiological, and functional evaluation following acute stroke. *British Journal of Clinical Pharmacology*. 1992;34(6):477-485. doi:10.1111/j.1365-2125.1992.tb05654.
20. Dyken ML. Precipitating Factors, Prognosis, and Demography of Cerebrovascular Disease in an Indiana Community: A Review of All Patients Hospitalized from 1963 to 1965 With Neurologica Examination of Survivors. *Stroke*. 1970;1(4):261-269. doi:10.1161/01.STR.1.4.261
21. Harvey RL, Edwards D, Dunning K, et al. Randomized Sham-Controlled Trial of Navigated Repetitive Transcranial Magnetic Stimulation for Motor Recovery in Stroke: The NICHE Trial. *Stroke*. 2018;49(9):2138-2146. doi:10.1161/STROKEAHA.117.020607
22. Du J, Tian L, Liu W, et al. Effects of repetitive transcranial magnetic stimulation on motor recovery and motor cortex excitability in patients with stroke: a randomized controlled trial. *Eur J Neurol*. 2016;23(11):1666-1672. doi:10.1111/ene.13105
23. Guan YZ, Li J, Zhang XW, et al. Effectiveness of repetitive transcranial magnetic stimulation (rTMS) after acute stroke: A one-year longitudinal randomized trial. *CNS Neurosci Ther*. 2017;23(12):940-946. doi:10.1111/cns.12
24. Kim WS, Kwon BS, Seo HG, Park J, Paik NJ. Low-Frequency Repetitive Transcranial Magnetic Stimulation Over Contralesional Motor Cortex for Motor Recovery in Subacute Ischemic Stroke: A Randomized Sham-Controlled Trial. *Neurorehabil Neural Repair*. 2020;34(9):856-867. doi:10.1177/1545968320 948610
25. Sasaki N, Mizutani S, Kakuda W, Abo M. Comparison of the Effects of High- and Low-frequency Repetitive Transcranial Magnetic Stimulation on Upper Limb Hemiparesis in the Early Phase of Stroke. *Journal of Stroke and Cerebrovascular Diseases*. 2013;22(4):413-418. doi:10.1016/j.jstrokecerebrovasdis.2011.10.004
26. Yang Y, Pan H, Pan W, et al. Repetitive Transcranial Magnetic Stimulation on the Affected Hemisphere Enhances Hand Functional Recovery in Subacute Adult Stroke Patients: A Randomized Trial. *Front Aging Neurosci*. 2021;13:636184. doi:10.3389/fnagi.2021.636184
27. Hosomi K, Morris S, Sakamoto T, et al. Daily Repetitive Transcranial Magnetic Stimulation for Poststroke Upper Limb Paresis in the Subacute Period. *Journal of Stroke and Cerebrovascular Diseases*. 2016;25(7):1655-1664. doi:https://doi.org/10.1016/j.jstrokecerebrovasdis.2016.02.024
28. Sharma H, Vishnu VY, Kumar N, et al. Efficacy of Low-Frequency Repetitive Transcranial Magnetic Stimulation in Ischemic Stroke: A Double-Blind Randomized Controlled Trial. *Archives of Rehabilitation Research and Clinical Translation*. 2020;2(1):100039. doi:10.1016/j.arrct.2020.100039
29. Barros Galvão SC, Borba Costa dos Santos R, Borba dos Santos P, Cabral ME, Monte-Silva K. Efficacy of Coupling Repetitive Transcranial Magnetic Stimulation and Physical Therapy to Reduce Upper-Limb Spasticity in Patients with Stroke: A Randomized Controlled Trial. *Archives of Physical Medicine and Rehabilitation*. 2014;95(2):222-229. doi:10.1016/j.apmr.2013.10.023

30. Bonin Pinto C, Morales-Quezada L, de Toledo Piza PV, et al. Combining Fluoxetine and rTMS in Poststroke Motor Recovery: A Placebo-Controlled Double-Blind Randomized Phase 2 Clinical Trial. *Neurorehabil Neural Repair*. 2019;33(8):643-655. doi:10.1177/1545968319860483
31. Wilson JTL, Hareendran A, Grant M, et al. Improving the Assessment of Outcomes in Stroke: Use of a Structured Interview to Assign Grades on the Modified Rankin Scale. *Stroke*. 2002;33(9):2243-2246. doi:10.1161/01.STR.0000027437.22450.BD
32. Quinn TJ, Dawson J, Walters MR, Lees KR. Exploring the Reliability of the Modified Rankin Scale. *Stroke*. 2009;40(3):762-766. doi:10.1161/STROKEAHA.108.522516
33. New PW, Buchbinder R. Critical Appraisal and Review of the Rankin Scale and Its Derivatives. *Neuroepidemiology*. 2006;26(1):4-15. doi:10.1159/000089536
34. Du J, Yang F, Hu J, et al. Effects of high- and low-frequency repetitive transcranial magnetic stimulation on motor recovery in early stroke patients: Evidence from a randomized controlled trial with clinical, neurophysiological and functional imaging assessments. *NeuroImage: Clinical*. 2019;21:101620. doi:10.1016/j.nicl.2018.101620
35. van der Lee JH, Beckerman H, Lankhorst GJ, Bouter LM. The Responsiveness of the Action Research Arm Test and the Fugl-Meyer Assessment Scale in Chronic Stroke Patients. *Journal of Rehabilitation Medicine*. 2001;33(3):110-113. doi:10.1080/165019701750165916
36. Wood-Dauphinee SL, Williams JI, Shapiro SH. Examining outcome measures in a clinical study of stroke. *Stroke*. 1990;21(5):731-739. doi:10.1161/01.STR.21.5.731
37. Lin JH, Hsueh IP, Sheu CF, Hsieh CL. Psychometric properties of the sensory scale of the Fugl-Meyer Assessment in stroke patients. *Clin Rehabil*. 2004;18(4):391-397. doi:10.1191/0269215504cr737oa
38. Bertrand AM, Fournier K, Wick Brasey MG, Kaiser ML, Frischknecht R, Diserens K. Reliability of maximal grip strength measurements and grip strength recovery following a stroke. *Journal of Hand Therapy*. 2015;28(4):356-363. doi:10.1016/j.jht.2015.04.004
39. Hsieh Y Wei, Wu C Yi, Liao W Wen, Lin K Chung, Wu K Yuh, Lee C Yi. Effects of Treatment Intensity in Upper Limb Robot-Assisted Therapy for Chronic Stroke: A Pilot Randomized Controlled Trial. *Neurorehabil Neural Repair*. 2011;25(6):503-511. doi:10.1177/1545968310394871
40. Mayer NH, Esquenazi A. Muscle overactivity and movement dysfunction in the upper motoneuron syndrome. *Physical Medicine and Rehabilitation Clinics of North America*. 2003;14(4):855-883. doi:10.1016/S1047-9651(03)00093-7
41. Duncan PW, Zorowitz R, Bates B, et al. Management of Adult Stroke Rehabilitation Care: A Clinical Practice Guideline. *Stroke*. 2005;36(9). doi:10.1161/01.STR.0000180861.54180.FF
42. Brown P. Pathophysiology of spasticity. *Journal of Neurology, Neurosurgery & Psychiatry*. 1994;57(7):773-777. doi:10.1136/jnnp.57.7.773
43. Urban PP, Wolf T, Uebele M, et al. Occurence and Clinical Predictors of Spasticity After Ischemic Stroke. *Stroke*. 2010;41(9):2016-2020. doi:10.1161/STROKEAHA.110.581991

44. Muus I, Petzold M, Ringsberg KC. Health-related quality of life among Danish patients 3 and 12 months after TIA or mild stroke: Health-related quality of life among Danish patients. *Scandinavian Journal of Caring Sciences*. 2010;24(2):211-218. doi:10.1111/j.1471-6712.2009.00705.x
45. Williams LS, Weinberger M, Harris LE, Clark DO, Biller J. Development of a Stroke-Specific Quality of Life Scale. *Stroke*. 1999;30(7):1362-1369. doi:10.1161/01.STR.30.7.1362
46. Butler AJ, Wolf SL. Putting the Brain on the Map: Use of Transcranial Magnetic Stimulation to Assess and Induce Cortical Plasticity of Upper-Extremity Movement. *Physical Therapy*. 2007;87(6):719-736. doi:10.2522/ptj.20060274
47. Siebner H, Rothwell J. Transcranial magnetic stimulation: new insights into representational cortical plasticity. *Experimental Brain Research*. 2003; 148(1):1-16.doi:10.1007/s00221-002-1234-2
48. Schambra HM, Sawaki L, Cohen LG. Modulation of excitability of human motor cortex (M1) by 1 Hz transcranial magnetic stimulation of the contralateral M1. *Clinical Neurophysiology*. 2003;114(1):130-133. doi:10.1016/S1388-2457(02)00342-5
49. Kuzu Ö, Adiguzel E, Kesikburun S, Yaşar E, Yılmaz B. The Effect of Sham Controlled Continuous Theta Burst Stimulation and Low Frequency Repetitive Transcranial Magnetic Stimulation on Upper Extremity Spasticity and Functional Recovery in Chronic Ischemic Stroke Patients. *Journal of Stroke and Cerebrovascular Diseases*. 2021;30(7):105795. doi:10.1016/j.jstrokecerebrovasdis.2021.105795
50. Liu X Bo, Zhong J Guo, Xiao X Li, et al. Theta burst stimulation for upper limb motor dysfunction in patients with stroke: A protocol of systematic review and meta-analysis. *Medicine*. 2019;98(46):e17929. doi:10.1097/MD.000000 0000017929
51. Chen YJ, Huang YZ, Chen CY, et al. Intermittent theta burst stimulation enhances upper limb motor function in patients with chronic stroke: a pilot randomized controlled trial. *BMC Neurol*. 2019;19(1):69. doi:10.1186/s12883-019-1302-x
52. Sung WH, Wang CP, Chou CL, Chen YC, Chang YC, Tsai PY. Efficacy of Coupling Inhibitory and Facilitatory Repetitive Transcranial Magnetic Stimulation to Enhance Motor Recovery in Hemiplegic Stroke Patients. *Stroke*. 2013;44(5):1375-1382. doi:10.1161/STROKEAHA.111.000522
53. Talelli P, Wallace A, Dileone M, et al. Theta Burst Stimulation in the Rehabilitation of the Upper Limb: A Semirandomized, Placebo-Controlled Trial in Chronic Stroke Patients. *Neurorehabil Neural Repair*. 2012;26(8):976-987. doi:10.1177/1545968312437940
54. Chen Y, Wei QC, Zhang MZ, et al. Cerebellar Intermittent Theta-Burst Stimulation Reduces Upper Limb Spasticity After Subacute Stroke: A Randomized Controlled Trial. *Front Neural Circuits*. 2021;15:655502. doi:10.3389/fncir.2021.655502
55. Chen YH, Chen CL, Huang YZ, et al. Augmented efficacy of intermittent theta burst stimulation on the virtual reality-based cycling training for upper limb function in patients with stroke: a double-blinded, randomized controlled trial. *J NeuroEngineering Rehabil*. 2021;18(1):91. doi:10.1186/s12984-021-00885-5

56. Meng Y, Zhang D, Hai H, Zhao YY, Ma YW. Efficacy of coupling intermittent theta-burst stimulation and 1 Hz repetitive transcranial magnetic stimulation to enhance upper limb motor recovery in subacute stroke patients: A randomized controlled trial. *RNN*. 2020;38(1):109-118. doi:10.3233/RNN-190953
57. Hsu YF, Huang YZ, Lin YY, et al. Intermittent theta burst stimulation over ipsilesional primary motor cortex of subacute ischemic stroke patients: A pilot study. *Brain Stimulation*. 2013;6(2):166-174. doi:10.1016/j.brs.2012.04.007
58. Ackerley SJ, Byblow WD, Barber PA, MacDonald H, McIntyre-Robinson A, Stinear CM. Primed Physical Therapy Enhances Recovery of Upper Limb Function in Chronic Stroke Patients. *Neurorehabil Neural Repair*. 2016;30(4):339-348. doi:10.1177/1545968315595285
59. Xiang H, Sun J, Tang X, Zeng K, Wu X. The effect and optimal parameters of repetitive transcranial magnetic stimulation on motor recovery in stroke patients: a systematic review and meta-analysis of randomized controlled trials. *Clin Rehabil*. 2019;33(5):847-864. doi:10.1177/0269215519829897
60. Ling H, Tao T, Xu J, Xu D. Effects of repetitive transcranial magnetic stimulation on upper limb motor function in patients with stroke: a meta analysis. *National Medical Journal of China*. 2017;97(47):3739-3745. doi:10.3760/ cma.j.issn.0376-2491.2017.47.012
61. Wassermann EM. Risk and safety of repetitive transcranial magnetic stimulation: report and suggested guidelines from the International Workshop on the Safety of Repetitive Transcranial Magnetic Stimulation, June 5–7, 1996. *Electroencephalography and Clinical Neurophysiology/ Evoked Potentials Section*. 1998;108(1):1-16. doi:10.1016/S0168-5597(97)00096-8
62. Zaghi S, Heine N, Fregni F. Brain stimulation for the treatment of pain: A review of costs, clinical effects, and mechanisms of treatment for three different central neuromodulatory approaches. Published online 2010:16.
63. Doris Miu KY, Kok C, Leung SS, Chan EYL, Wong E. Comparison of Repetitive Transcranial Magnetic Stimulation and Transcranial Direct Current Stimulation on Upper Limb Recovery Among Patients with Recent Stroke. *Ann Rehabil Med*. 2020;44(6):428-437. doi:10.5535/arm.20093
64. Felipe Fregni CP. A Combined Therapeutic Approach in Stroke Rehabilitation: A Review on Non-Invasive Brain Stimulation plus Pharmacotherapy. *Int J Neurorehabilitation Eng*. 2014;01(03). doi:10.4172/ 2376-0281.1000123
65. Alonso-Alonso M, Fregni F, Pascual-Leone A. Brain Stimulation in Poststroke Rehabilitation. *Cerebrovasc Dis*. 2007;24(Suppl. 1):157-166. doi:10.1159/000107392
66. Nicolo P, Magnin C, Pedrazzini E, et al. Comparison of Neuroplastic Responses to Cathodal Transcranial Direct Current Stimulation and Continuous Theta Burst Stimulation in Subacute Stroke. *Archives of Physical Medicine and Rehabilitation*. 2018;99(5):862-872.e1. doi:10.1016/j.apmr.2017.10.026
67. Di Lazzaro V, Rothwell JC. Corticospinal activity evoked and modulated by non-invasive stimulation of the intact human motor cortex. *J Physiol*. 2014;592(19):4115-4128. doi:10.1113/jphysiol.2014.274316
68. Bikson M, Rahman A. Origins of specificity during tDCS: anatomical, activity-

selective, and input-bias mechanisms. *Front Hum Neurosci*. 2013;7. doi:10.3389/fnhum.2013.00688

69. Health C for D and R. Repetitive Transcranial Magnetic Stimulation (rTMS) Systems - Class II Special Controls Guidance for Industry and FDA Staff. *FDA*. Published online March 25, 2021. Accessed July 4, 2022. https://www.fda.gov/medical-devices/guidance-documents-medical-devices-and-radiation-emitting-products/repetitive-transcranial-magnetic-stimulation-rtms-systems-class-ii-special-controls-guidance
70. Lefebvre S, Thonnard JL, Laloux P, Peeters A, Jamart J, Vandermeeren Y. Single Session of Dual-tDCS Transiently Improves Precision Grip and Dexterity of the Paretic Hand After Stroke. *Neurorehabil Neural Repair*. 2014;28(2):100-110. doi:10.1177/1545968313478485
71. Dionísio A, Duarte IC, Patrício M, Castelo-Branco M. The Use of Repetitive Transcranial Magnetic Stimulation for Stroke Rehabilitation: A Systematic Review. *Journal of Stroke and Cerebrovascular Diseases*. 2018;27(1):1-31. doi:10.1016/j.jstrokecerebrovasdis.2017.09.008
72. Di Pino G, Pellegrino G, Assenza G, et al. Modulation of brain plasticity in stroke: a novel model for neurorehabilitation. *Nat Rev Neurol*. 2014;10 (10):597-608. doi:10.1038/nrneurol.2014.162

The Influence of Injunctive Social Norms Appeal on Behavioral Intention to Participate in Blood Donation

SaraVotey Mom

Author Background: *SaraVotey Mom grew up in Cambodia and currently attends Liger Leadership Academy in Phnom Penh, Cambodia. Her Pioneer research concentration was in the field of business and titled "Persuasive Marketing."*

Abstract

Offering to participate in blood donations is beyond a precious gift. It is also essential to ensure that there is a sufficient number of donations to make the blood supply system run. Unfortunately, blood shortage is becoming a global crisis, and blood transfusion centers are encountering challenges in designing relevant and impactful strategies to motivate voluntary blood donations. For this reason, it is requisite to research how the uses of certain sophisticated and appropriate marketing communication strategies contribute to encouraging participation in donations. The purpose of this study is to examine how the injunctive social norm marketing technique influences people's behavioral intention to donate blood by randomly assigning the participants to one of the two crafted advertisements. One includes an experimentally manipulated message about injunctive norms, and one includes a control generic message about the imperative need for blood donation. The results suggest that injunctive normative messages do not contribute to boosting the behavioral intention of individuals involved in blood donation behaviors, but they indicate that people are influenced to donate through social norms in general. This paper concludes with a discussion about the reasons behind the findings, recommendations for further research, as well as the implications of the results. This is the first research looking at the effectiveness of injunctive norms on blood donation-related behaviors, making a theoretical contribution to the creation of further studies about social norms appeals and blood donation.

1. Introduction

Blood donation plays a significant role in the healthcare system worldwide. It is indispensable through its contribution to increasing the life expectancy of individuals with chronic and acute illnesses, saving millions of lives annually, as well as advancing medical intervention and surgical procedures in various ways. This research aims to investigate how a persuasive message influences individuals' behavioral intention toward voluntary blood donation.

1.1. Blood Donation

Every region of the world faces pressing issues in obtaining sufficient blood from safe donors to meet the nation's demands. While systems based on replacement donation by family members of the patients are hardly able to respond to clinical demands for blood, and paid donation constitutes a hazard to donors' and recipients' health, non-remunerated blood transfusion from donors is recognized by the World Health Organization as exceptionally important for the safety, availability, and sustainability of national blood supplies (WHO, 2010). Recruiting a number of safe, low-risk donors is an emerging challenge. In particular, recruitment in developing countries with barriers from the inadequacy of technical expertise has the potential to impede the efforts to achieve a truly voluntary blood donation system (Zaller et al., 2005). Based on WHO's study, 75 countries reported less than 10 donations of blood per 1000 people, while 72 countries still collected more than 50% of their blood supply from both replacement and paid donors (WHO, 2015).

In response to the COVID-19 pandemic, the initiation of preventive measures imposed negative health, social and economic impacts, including on the blood supply system (Quee et al., 2022). Due to cancellations of blood drives, an increase in social distancing, and personal protective equipment requirements during the pandemic, blood collections were hindered, worsening the issue of blood shortages in many parts of the world (Murphy et al., 2021). The decline in blood donation does not mean that the blood demand in the surgical and intensive care areas has declined, which has prompted many blood collection institutions to emphasize the ongoing requirements for participation in blood donation.

The shortage of supply against demand has had influences not only on individuals who experience serious injuries or undergo clinical surgery, but also had great repercussions on patients who need regular blood transfusion therapy, including those suffering from thalassemia, nutritional anemia, leukemia, or kidney disease (Arshad Ali et al., 2021). For instance, thalassemia is a hereditary blood disorder in which the body makes an abnormal form or inadequate amount of hemoglobin. It is a disorder of hemoglobin synthesis that requires a regular blood transfusion (Gomes, 2021). A great number of thalassemia patients depend on blood transfusions to thrive, which indicates that the lack of involvement in donating blood puts their lives in jeopardy. Additionally, the rates of maternal mortality can also be minimized through blood transfusion. If a woman is unwell or anemic, she may require a blood transfusion to restore her hemoglobin level. On certain occasions, blood is kept ready before risky

circumstances arise. Because many women die while going through parturition, blood transfusion plays an integral and requisite role in obstetric practice (Kathpalia et al., 2016). Blood crisis undeniably signifies that there is also a crisis in blood transfusion, which further jeopardizes the well-being of mothers.

Because maintaining an adequate supply of blood enhances the well-being of the population and saves lives, the failure to acquire blood donations is regarded as both a national and international security threat. In fact, blood shortages would unfavorably impact blood transfusion services and result in the collapse of the health system and health security. Consequently, the downfall in health security also gives the possibility of the downfall in social security and political security (Ibrahim, 2021).

In a period where demand for blood donation is increasing while the number of donors is decreasing, it is fundamental to utilize sophisticated types of communication that appear to hold considerable potential to influence and facilitate voluntary blood donation. Blood facilities are facing a profound difficulty when it comes to getting new donors and keeping regular donors loyal (Mews & Boenigk, 2012). This firmly demonstrates that there is a necessity to research and execute highly effective messaging and persuasive marketing strategies that could influence the population's behavior toward blood donation. As the blood crisis is a global issue, individuals who meet the eligibility guidelines to donate including those who are in good health, aged between 18 and 65, and weigh at least 50 kg should be targeted with these persuasive marketing strategies (WHO, 2020).

Common measures taken to sustain the blood supplies include raising awareness and motivating people through accessible blood donation centers and social media. However, to meet the requirements of blood supply, more effective message framing methods have to be examined. To increase the donor pool by motivating the population for voluntary blood donation, this paper aims to address the matter through the examination of the usage of persuasive messaging that attempts to alter people's behavioral intention toward voluntary blood donation. This study attempted to craft an advertisement using a persuasive technique and test the effectiveness of the message within an experiment toward changing people's behavior toward blood donation.

1.2. Persuasive messaging

The art of persuasion is known to be a potent force that shifts people's decisions and actions. It is also described as an interactive process in which a message has the power to influence an individual's understanding, knowledge, or beliefs. Focusing on communicating a certain message to the receivers, persuasion attempts to achieve its goal through reasonable and sensible expressions (Lee & Xia, 2011).

It is widely recognized that persuasive messages tend to stimulate attitude, intention, and behavior change. Researchers have noted that how the messages are framed completely shape the effectiveness of persuasion (Carfora & Catellani, 2021). For example, messages can be framed in terms of the positive or negative consequences, that is, regarding the loss or gain valence of a given behavior. The implication of prospect theory implies that framed and factually equivalent messages affect individuals in different ways by emphasizing the

benefits or costs of committing a certain behavior. The loss-framed messages feature failing to attain, avoiding a desirable outcome, and attaining an undesirable outcome as the consequence of the given behavior. In contrast, gain-framed messages emphasize attaining a desirable outcome or great results. In the context of cessation of smoking, "Do not quit smoking and you will die sooner" is a loss-framed message, and "Quit smoking and you will live long" is a gain-framed message (Toll et al., 2007). A particular relevance to the theory of gains and losses framing is loss aversion. Loss aversion is the conception that individuals tend to gravitate toward disliking losses more than equivalent gains (Kahneman & Tversky, 1979). Empirical evidence suggests that loss frames have a stronger influence, recommending that negative communication has stronger supremacy on the adoption of decisions than equivalent positive communication. This setting of message framing has been enforced and emphasized under the interventions of both health and economic consequences (Hameleers & Boukes, 2021).

Changing people's behavior through persuasive message framing is prominent as a form of solving social problems. It involves changing individuals' habits by transforming their harmful behaviors into more productive behaviors, changing society's values and beliefs, as well as creating social technology in order to increase the community's quality of life (Fatmawati, 2010). Accordingly, the research of persuasive message design is central through its contribution to designing and shaping subsequent campaign or intervention messages. Beyond changing attitudes and improving behaviors, the presence of research on this dynamic could suggest the different types of messages needed for different audiences and identify how alternative possible persuasive messages are likely to be relatively more effective in behavioral outcomes among consumers (O'Keefe, 2021).

1.3. Social Norms

Social norms are regarded to be essential components of the revolution of human behaviors. In the realm of a persuasive message, it is suggested that "humans are especially motivated to understand and to follow the norms of groups that we belong to and care about" (Tankard & Paluck, 2016), as it is our tendency to bring our thoughts and intentions in line with the desired perception in society. Persuasive interventions based on social norms are efficacious in reinforcing collective changes in behavior, as it is evidenced that "researchers and practitioners attempt to transform behavior in order to increase environmental and social sustainability in real-world contexts." Various initiatives of social norms have been enforced to broaden sustainability both environmentally and socially in the issues of energy, transportation, food, hygiene, charity, and alcohol consumption (Yamin et al., 2019).

The research on persuasive messages indicates that behaviors are transformed by two types of social norms: injunctive norms and descriptive norms. An injunctive norm corresponds to that of which most people approve or disapprove. It is commonly reflected in the expectation of the community or individuals who are important to the subject about the behaviors for them to adopt. Generally, injunctive norm messages exert a form of social pressure that advise what individuals must do or what is typically expected of them (Trelohan,

2021) (e.g., “Most of the blood donors disapprove of not donating blood even once in a lifetime”). In contrast, a descriptive norm corresponds to what most people do (e.g., an estimated 6.8 million people in the U.S. donate blood (American Red Cross, 2018), or more than half of Saudis said they frequently donate their blood (McCarthy, 2018)). It is also known as the “standards of behavior adopted by the majority of people” because humans have the inclination to take into account their surroundings and attune their behavior to conform (Trelohan, 2021).

2. Literature Review

Previous research was conducted by Slaunwhite et al. (2009) using the injunctive/descriptive framework to investigate how social norm appeals impact stair climbing behavior based on the principles of the Focus Theory of Normative Conduct. In the research, four different posters of varying perspectives on norm-based conditions were presented, including one injunctive norm-based appeal, one descriptive norm-based appeal, one that integrated both injunctive and descriptive in a consistent manner (promoting that stair use was both good and common), and one that integrated both injunctive and descriptive in an inconsistent manner (promoting that stair use was good but not common). After comparing the effectiveness of these norm-based appeals to the standard messages recommended by organizations with logistic regression analyses, the results suggested that “stair climbing behavior rose considerably in the injunctive (+5.9%) and injunctive/descriptive consistent (+5.4%) poster conditions over the control conditions. It is recommended that messages originating from a norm-based framework are more significant for practitioners attempting to promote healthy behavior at work compared to generic information-based posters (Slaunwhite et al., 2009).

Another attempt at scrutinizing the effectiveness of appeals employing social norms was a study done by Goldstein et al. (2008). This study examined the effectiveness of an appeal that conveys the descriptive norm of participation in environmental conservation programs at encouraging towel reuse in hotels compared to the conventional appeals in the industry. In the experiment, two kinds of messages were designed. One message demonstrated the standard approach, centering on the importance of protecting the environment through their actions toward the hotel guests. The other message provided explicit descriptive norms to the guest, illustrating that the majority of other guests participate in the towel reuse program at least once in their stay at the hotel. Its message was "JOIN YOUR FELLOW GUESTS IN HELPING TO SAVE THE ENVIRONMENT. Almost 75% of guests who are asked to participate in our new resource savings program do help by using their towels more than once. You can join your fellow guests in this program to help save the environment by reusing your towels during your stay." Based on the analysis, the study showed that the normative message yielded a towel reuse rate that was significantly higher than the standard approach message, which contributed to illustrating the persuasiveness of descriptive norms to encourage individuals to engage in the real-world domain of environmental conservation (Goldstein et al., 2008).

Although the foregoing studies have shown implications and have

incorporated the theoretical framework of descriptive and injunctive social norms into the implementation of health and environment-related issues, it is crucial to note that each study has different target behaviors, contexts, and populations. The usage of social norm persuasive messages does not always have the possibility to impact every aspect and behavior of health or environment areas in the same fashion.

3. Overview

In the arena of donation and descriptive social norms, previous literature demonstrated that individuals have a great tendency to contribute to creating a positive impact for the public if others also do so. The study by XIE et al. (2019) establishes that descriptive norms promoted participants' willingness to donate blood voluntarily, but did not promote their actual donation of blood (XIE et al., 2019).

Because blood donation is factually not the behavior adopted by the majority of the citizens, I suggest that using injunctive norms as opposed to descriptive norms has the potential to influence the behavioral intentions among individuals to donate blood. Therefore, I hypothesize:

Hypothesis 1: Participants who view an advertisement with an injunctive norm message will be more likely to donate blood/have more positive attitudes toward donating blood than those who are in the control group.

To test this hypothesis, an experiment is carried out with two conditions: an injunctive norm condition and a control condition. Participants then report their attitudes and behavioral intentions based on the advertisement. This work of research contributes to previous research by examining the relationship between injunctive social norms and individuals' motives to donate blood.

3.1. Methodology

An experiment was conducted to address my hypothesis. 105 participants were recruited from mTurk. Participants were randomly assigned to one of two conditions. In the experimental condition, participants were presented with a blood donation advertisement with the main description to examine the influence of injunctive social norms on their behavioral intention to donate blood: "9 out of 10 people believe that donating blood is the right thing to do." In the control condition, participants were presented with an advertisement that was comprised of a general fact about the imperative need of participating in donating blood with the main description: "More than 38000 blood donations are needed every day." Both the experimental and control conditions included identical headings (which stated "Donate blood, Save Life"), identical pictures, and identical bottom descriptions (which stated "Register Now!" and "Your help matters"). The contrast in the main descriptions of the advertisement was the only difference between experimental and control conditions throughout the entirety of the study. (See Appendix A for injunctive norm advertisement, See Appendix B for control advertisement)

After viewing the advertisement, participants reflected on their general thoughts about what the experiment was about, what they liked about it, and the part of the experiment that stood out the most to them. They then reported the degree of their agreement or disagreement on a series of seven statements about the impact of these advertisements on their incentives to register and encourage others to donate, as well as their emotions of sadness and guilt after seeing the advertisement on a scale of 1 - 5 (strongly disagree to strongly agree). These were the statements: *I like this advertisement; This advertisement motivates me to donate blood; After viewing this advertisement, I would tell others to donate blood; The advertisement makes me want to register to donate; The advertisement makes me feel guilty; The advertisement makes me feel sad for those who need blood; The advertisement makes me want to talk to my family and friends to go donate together.*

Noting that each participant only saw one of the two advertisements, participants were then asked about their likelihood to contribute to donating blood in the next 3 months after seeing that specific advertisement on a scale of 1 - 5 (extremely unlikely to extremely likely).

Those who were in the experimental condition were then questioned on the effectiveness of the main description of their advertisement "9 out of 10 people believe that donating blood is the right thing to do" to their motivation to participate by rating on a scale from 1 to 10 (1 being not impacted and 10 being very much impacted). Those who were in the control condition were also questioned on the effectiveness of the main description of their advertisement "More than 38000 blood donations are needed every day," by rating on the same scale from 1 to 10. All participants were also asked a closed question about whether they believe that donating blood is the right thing to do with answer choices of Yes, No, and Maybe.

To understand more about the current behavioral intentions to donate of all participants, this study went on to ask participants about their overall experience with blood donation. They were asked closed-ended and open questions on whether they are a donor and the reasons they choose or choose not to donate blood. If they are a donor, they were asked to choose from a multiple-choice question with the following options: *I see people around me also donate; I have been asked; I donate for my own health benefits.* If they are not a donor, they are asked to elaborate on their choice. At the same time, they were also asked to rate the extent to which they agreed with the following statements on a scale of 1 - 5 (strongly disagree to strongly agree): *I believe that donating blood is ethical; Most people that are important to me would appreciate if I am a donor; If my colleagues or friends donated blood, I am likely to donate too; If my loved ones are donors and value donation, I am likely to donate blood.*

Finally, participants were requested to indicate their demographic information including their country, their gender, and their age range.

3.2. Results

Sample size = 10

3.2.1. Demographics

40% of the participants were female and 60% were male. Participants' ages varied widely with 24.8% aged 18-28, 45.7% aged 29-38, 19% aged 39-48, 8.6% aged 49-58, and 1.9% aged over 58 years old. Participants were from all over the world, with most being from the United States. 89% of participants have donated blood.

3.2.2. The Effectiveness of Message Framing

To test my hypotheses, I ran several ANOVAs comparing the control group's response to the experimental (i.e., social norms) group's response based on the advertisement. See Table 1 for the results. The groups did not differ significantly for the dependent variables; however, they were trending such that the control advertisement was more motivating and effective than the experimental group except for the item "this advertisement makes me feel sad for those who need blood," which had a higher mean in the experimental group.

Table 1.

Dependent Variable	Control Group Mean	Experimental Group Mean (Social Norms)	F-value	P-Value
I like this advertisement	4.02	3.87	.65	.42
This advertisement motivates me to donate blood	4.00	3.96	.04	.85
After viewing this advertisement, I would tell others to donate blood	4.14	3.95	1.02	.31
The advertisement makes me want to register to donate	3.90	3.71	.99	.32
The advertisement makes me feel guilty	3.10	3.33	.88	.35
This advertisement makes me feel sad for those who need blood	3.84	3.40	3.3	.07
The advertisement makes me want to talk to my family and friends to go donate together	3.86	3.76	.27	.61
How likely will you contribute to donate blood in the next 3 months after seeing the advertisement?	4.12	4.05	.11	.74

A thematic qualitative analysis was conducted on questions related to the advertisement including "What do you like about the advertisement" and "What part of the advertisement stood out to you the most?" The following themes emerged:

- The concept of the advertisement is informative and easy to follow through its pleasant colors and simple phrases including "Donate Blood, Save Life"
- The main statistics of both advertisements which stated "9 out of 10 people believe that donating blood is the right thing to do" and "More than 38000 blood donations are needed every day" stood out most to the participants with the same degree.

3.2.3. Understanding Blood Donation Behavior

Beliefs about blood donation were measured using four scale questions and one categorical question. See Tables 2, 3, and 4 below. Overall, people believe that it is crucial to donate and think that it is the right thing to do. The majority of the participants are more likely influenced to donate if their family or colleagues are donors.

Table 2.

Variables	Mean	SD
I believe that donating blood is ethical	4.02	.945
Most people that are important to me would appreciate if I am a donor	3.90	.909
If my colleagues or friends donated blood, I am likely to donate too	4.15	.911
If my loved ones are donors and value donation, I am likely to donate blood	4.04	.835

Table 3. *Is Donating the Right Thing to Do?*

Variables	Number	Percent
Yes	97	92.4%
No	2	1.9%
Maybe	6	5.7%

Table 4. *Why do you donate?*

Variables	Number	Percent
People around me donate	53	54.6%
I have been asked	23	12.7%
I donate for my own health benefits	21	21.6%

A thematic qualitative analysis was conducted on questions related to why participants do not donate blood. Themes include:

- Having health issues
- Never gotten the opportunity to donate
- Lack of awareness about the procedure and places to donate
- Afraid of needles

To address the effectiveness of various marketing messages, participants both in the experimental condition and the control condition were asked about the effectiveness of the main description of the advertisement they see on the motivation to participate in blood donation on a scale of 1 to 10. See Table 5 below. Overall, both the main description of social norms and control variables have almost equivalent influence on people's intention to donate blood, having the same mean with a slightly different standard deviation.

Table 5.

Variables	Mean	SD
How does the phrase "9 out of 10 people believe that donating blood is the right thing to do" impact you to participate in blood donation? Rate on a scale of 1 to 10 (1 = Not impacted, 10 = very much impacted)	8.0000	1.90332
How does the phrase "More than 38000 blood donations are needed every day" impact you to participate in blood donation? Rate on a scale of 1 to 10 (1 = Not impacted, 10 = very much impacted)	8.0000	1.90595

4. Discussion

Acknowledging the necessity of empirical research endeavors in the literature on persuasive marketing, consumer behavior, and features that affected consumer-oriented prosocial behavior and especially the lack of research promoting the propensity of blood donation, practitioners and researchers accentuate the importance of more research in the areas (Goldstein et al., 2008). The objective of this research was to incorporate a theoretical framework of a type of persuasive messaging which is the injunctive social norm and its influence on individuals' behavioral intention toward voluntary blood donation.

The study demonstrated in this paper illustrates that the usage of the injunctive social norm has no significant influence to motivate the population to engage in the real-world issue of blood donation. With respect to the generally higher mean score for the control message, it is suggested that the main description of the injunctive social norms advertisement which stated "9 out of 10 people believe that donating blood is the right thing to do" has less power to

motivate people to donate blood, make them tell other to donate blood, make them register, make them feel sad, and make them likely to donate blood within the next 3 months (except making people feel guilty) than the main description of the control advertisement which comprised of a general fact about the imperative need of participating in donating blood which stated, "More than 38000 blood donations are needed every day". Therefore, after running the ANOVAs comparing the participants' degree of agreement or disagreement and recognizing the superiority of the control message advertisement, it shows that injunctive normative message and information is not an optimal approach compared to the standard control message to activate people in the domain and target market of blood donation.

This result is not consistent with our hypothesis, as it appears that participants who view an advertisement with an injunctive norm message were not more likely to donate blood/have more positive attitudes toward donating blood than those who are in the control group. However, our result plays a role in confirming that the social norm mechanism, in general, has an association with driving motivation and incentive for the behaviors of consumers in the setting of blood donation. As evidenced by the social norms-related phrases in Table 2, the mean scores of the phrases when the participant rated the extent they agreed on the phrase on a scale of 1-5 was 4.15 (for the social norm phrase: If my colleagues or friends donated blood, I am likely to donate too) and 4.04 (for the social norm phrase: If my loved ones are donors and value donation, I am likely to donate blood), strongly signifying that consumers' behaviors are influenced by their family and colleagues.

This study identified that injunctive social norms are not powerful in shaping the behavior of blood donation. This finding is analogous to the past finding examining the role of the injunctive norms on the COVID-19 vaccine intentions, suggesting that using injunctive norms in the formation of public opinions toward supporting the vaccination has a minimal effect on the intended COVID vaccine uptake (Carey et al., 2022). At the same time, this finding also contradicts other previous research on the behavior of injunctive norms. For instance, researchers from past work that focus on the relation of injunctive norms regarding economic consequences found that injunctive norms have a major impact on managers' budget reporting honesty (Altenburger, 2017).

The result of this study may differ for a variety of reasons, which could be by virtue of the nature of the behaviors studied and the relationship of the behavior's context toward injunctive norms. A possible explanation is that the blood donation behavior that is being studied in this paper is affected differently by persuasive messaging compared to other behaviors previously explored. Previous research involved motivating the population to undertake an activity, action, or attitude that the population thought/believed would not impose a high level of risk, harm, or worry: encouraging stair climbing behavior, and encouraging participation in environmental conservation programs by reusing towels in the hotel (Slaunwhite et al., 2009; Goldstein et al., 2008). However, this study involves a target behavior that is widely known to be difficult and complex to encourage participation. Society's perception and the potential deterrents to blood donation also impacted individuals' responses when being asked and encouraged to volunteer to donate. As demonstrated by our thematic qualitative analysis of the reasons why the participants don't donate blood, the pain from

phlebotomy, the concern about infection, and the misperception about the donation's eligibility/procedure all contributed to making this become a harder behavior to nudge. Accordingly, this confirms that the nature of the behavior chosen is a determining factor that impacts the effectiveness of the message. Because blood donation is an intricate behavior, this means that further research must continue to investigate how to productively utilize sophisticated types of messages to influence this behavior.

Another explanation could be the framing of the injunctive message of this study. Carey et al.'s research (2022) argued that when a large number of participants had relatively accurate perceptions of the injunctive norms, the potential effectiveness of the treatment can be restrained. This demonstrates that the potency of the main description of the injunctive norms advertisement is being minimized since the participants are already aware that the population greatly appreciates the act of donating blood and regards it as moral before the advertisement; therefore, they are not influenced by the normative message as hypothesized. Instead of framing the majority's belief that blood donation is the right thing to do, it is fundamental to further assess the influence of the social norm message by framing it in the context of approval or disapproval. On the other hand, the bias of the participants' responses toward the framing of the message could also be a determinant. Although humans have the tendency to conform to the majority and are influenced by social norms, sometimes they dislike admitting that they do conform and are prone to answering the questions in the experiment far beyond reality. It could be possible that because the participants are aware that the advertisement uses injunctive norms and do not want to be seen relying on what others believe to guide their behavior, they are less likely to accept expressing the influence of the normative message on their behavioral intentions. This strongly advocates for the significance of attentively framing the message to minimize the bias of the participant's response.

5. Recommendations and Implications

This research indicates that more studies need to be conducted to further explore how to successfully use persuasive messaging to change behaviors in blood donations. Silva and John's study (2017) has advocated that the convention of social norms is more convincing when being practiced in a stable and homogeneous population. Because the "effectiveness of the normative interventions is determined by the population having a shared sense of what is the desired form of that behavior" (Silva & John, 2017), we possibly could further investigate how impactful social norms and specifically injunctive norms are when being employed in a homogeneous or small environment where there is a shared identity among the people in the group. In particular, a study looking at how social norms and public good messages enhance tax compliance evidenced that minority norm statement produces a greater treatment effect than country norm statements, and that using the normative statements of late tax payment at the local level is more fruitful in increasing the payment rate than using the norm at the country level (Hallsworth et al., 2014). The samples of participants in this study are individuals from all over America, India, and Brazil, which is a heterogeneous population sharing different beliefs and behaviors toward blood

donation; therefore, their behaviors are not influenced by the norm as they don't identify with the wide group. Therefore, this study recommends future researchers frame the normative information in a smaller scope of the community, such as a regional/community norm instead of country/worldwide norm.

In 2019, a WRAP campaign called Recycle NOW used descriptive norm messaging "Everybody Does" to encourage people to recycle more (Gould, 2022). Recognizing the status of descriptive norm messaging within this earlier work, it is wise for behavioral researchers to carry on the research of the potential of not only injunctive norms like this study but also of descriptive norms. Marketers and researchers could attempt to manipulate descriptive norms to motivate behavioral intention of blood donations in locations with relatively high blood donation rates, which show that the majority are blood donors. This type of research could be done on small scales such as in institutions, organizations, or communes with existing high blood donation participation rates.

6. Limitations

Similar to other research, this current research also consisted of some limitations that could be categorized as lessons to apply to the implementation of future research. This research involved examining how persuasive marketing advertisements impact the behavioral intentions of the participants in terms of their interest in registering to donate blood. Because this study measured behavioral intention as opposed to real behavior, its results are not a complete representation to conclude that the persuasive message contributes to making people donate and tackle the issue of blood shortage. In short, it could only characterize how these messages evoke the willingness and motivation for people to donate blood. The study of XIE et al. (2019), where descriptive norms are effective in motivating people to donate blood but not promoting their actual donation, is an example portraying where studying only the behavioral intention could lead. Furthermore, as concisely mentioned in the discussion section, online participants are prone to many biases in terms of their responses to the experiment's questions. Although participants were randomly assigned to only one of two types of advertisement, there are still possibilities of bias. Blood donation could be viewed as a sensitive topic and this could significantly affect the responses of our participants. They may not tell the truth and may want to provide answers that lead in different directions. Prospective studies should consider larger sample sizes and/or bigger monetary incentives to overcome this limitation.

7. Conclusion

Although this study hints at the relative efficacy of the social norm persuasive messaging method in influencing the behavioral intention of voluntary blood donation, it suggests that an injunctive normative message is not a potent method to activate people's intentions in the behavior of blood donation. Nevertheless, its result is still theoretically and practically significant for contributing to future research intending to investigate similar persuasive messaging techniques or

target market behavior of blood donation. This study adds to the field of blood donation behavior and the discipline of marketing by emphasizing the primary function of further research looking at the effectiveness of injunctive norms or social norms to persuade people to donate blood or implement a course of desirable actions that create positive impacts on our societies. Further investigating the effectiveness of persuasive messaging is necessary to build a future where blood is available to save lives wherever and whenever it is needed.

References

Altenburger, M. (2017). The Effect of Injunctive Social Norms and Dissent on Budget Reporting Honesty. *Journal of International Accounting Research*, 16(2), 9–31. https://doi.org/10.2308/jiar-51744

Arshad Ali, S., Azim, D., Hassan, H., Iqbal, A., Ahmed, N., Kumar, S., & Nasim, S. (2021). The impact of COVID-19 on transfusion-dependent thalassemia patients of Karachi, Pakistan: A single-center experience. *Transfusion Clinique et Biologique*, 28(1), 60–67. https://doi.org/10.1016/j.tracli.2020.10.006

Carfora, V., & Catellani, P. (2021). The Effect of Persuasive Messages in Promoting Home-Based Physical Activity During COVID-19 Pandemic. *Frontiers in Psychology*, https://doi.org/10.3389/fpsyg.2021.644050

Carey, J. M., Keirns, T., Loewen, P. J., Merkley, E., Nyhan, B., Phillips, J. B., Rees, J. R., & Reifler, J. (2022). Minimal effects from injunctive norm and contentiousness treatments on COVID-19 vaccine intentions: evidence from 3 countries. *PNAS Nexus*, 1(2). https://doi.org/10.1093/pnasnexus/pgac031

Fatmawati, I. (2010). Changing Public Attitude and Behavior Through Persuasive Message Framing. *Academia*. https://www.academia.edu/8882983/Changing_Public_Attitude_and_Behavior_Through_Persuasive_Message_Framing

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels. *Journal of Consumer Research*, 35(3), 472–482. https://doi.org/10.1086/586910

Gomes, R. R. (2021). Hypogonadotropic Hypogonadism in a Female Patient with Thalassemia Major. *International Journal of Blood Research and Disorders*, 8(2). https://doi.org/10.23937/2469-5696/1410066

Gould, Tom. "The Power of Social Norms to Encourage Recycling." *Impact*, 29 Mar. 2022, https://impactmr.com/2022/03/29/the-power-of-social-norms-to-encourage-recycling/

Hallsworth, Michael, et al. "The Behavioralist As Tax Collector: Using Natural Field Experiments to Enhance Tax Compliance." *Journal of Public Economics*, 2014. *Crossref*, https://doi.org/10.3386/w20007

Hameleers, M., & Boukes, M. (2021). The Effect of Gain-versus-Loss Framing of Economic and Health Prospects of Different COVID-19 Interventions: An Experiment Integrating Equivalence and Emphasis Framing. *International Journal of Public Opinion Research*, 33(4), 927–945. https://doi.org/10.1093/ijpor/edab027

Ibrahim Omer Yahia, A. (2021). Management of Blood Supply and Blood Demand to Ensure International Health Security. Contemporary Developments and Perspectives in International Health Security - Volume 2. https://doi.org/10.5772/intechopen.96128

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263–291. https://doi.org/10.2307/1914185

Kathpalia, S., Chawla, J., Harith, A., Gupta, P., & Anveshi, A. (2016). Blood transfusion practices among delivery cases: A retrospective study of two years. *Medical Journal Armed Forces India*, 72, S43–S45. https://doi.org/10.1016/j.mjafi.2016.01.010

Lee, G., & Xia, W. (2011). A longitudinal experimental study on the interaction effects of persuasion quality, user training, and first-hand use on user perceptions of new information technology. *Information & Management*, 48(7), 288–295. https://doi.org/10.1016/j.im.2011.09.003

McCarthy, N. (2018, July 27). Where People Are Most Willing to Donate Blood. *Statista Infographics*. https://www.statista.com/chart/14892/where-people-are-most-willing-to-donate-blood/

Mews, M., & Boenigk, S. (2012). Does organizational reputation influence the willingness to donate blood? *International Review on Public and Nonprofit Marketing*, 10(1), 49–64. https://doi.org/10.1007/s12208-012-0090-4

Murphy, C., Fontaine, M., Luethy, P., McGann, H., & Jackson, B. (2021). Blood usage at a large academic center in Maryland in relation to the COVID-19 pandemic in 2020. *Transfusion*, 61(7), 2075–2081. https://doi.org/10.1111/trf.16415

O'Keefe, D. J. (2021). Persuasive Message Pretesting Using Non-Behavioral Outcomes: Differences in Attitudinal and Intention Effects as Diagnostic of Differences in Behavoral Effects. *Journal of Communication*, 71(4), 623–645. https://doi.org/10.1093/joc/jqab017

Quee, F. A., Spekman, M. L. C., Prinsze, F. J., Ramondt, S., Huis In 't Veld, E. M. J., Hurk, K., & Merz, E. (2022). Blood donor motivators during the COVID-19 pandemic. *Journal of Philanthropy and Marketing*. https://doi.org/10.1002/nvsm.1757

Silva, Antonio, and Peter John. "Social Norms Don't Always Work: An Experiment to Encourage More Efficient Fees Collection for Students." PLOS ONE, edited by Pablo Brañas-Garza, vol. 12, no. 5, 2017, p. e0177354. *Crossref*, https://doi.org/10.1371/journal.pone.0177354.

Slaunwhite, J.M., Smith, S.M., Fleming, M.T. and Fabrigar, L.R. (2009), "Using normative messages to increase healthy behaviours", *International Journal of Workplace Health Management*, Vol. 2 No. 3, pp. 231-244. https://doi.org/10.1108/17538350910993421

Tankard, M. E., & Paluck, E. L. (2016). Norm Perception as a Vehicle for Social Change. *Social Issues and Policy Review*, 10(1), 181–211. https://doi.org/10.1111/sipr.12022

Toll, B. A., O'Malley, S. S., Katulak, N. A., Wu, R., Dubin, J. A., Latimer, A., Meandzija, B., George, T. P., Jatlow, P., Cooney, J. L., & Salovey, P. (2007). Comparing gain- and loss-framed messages for smoking cessation with sustained-release bupropion: A randomized controlled trial. *Psychology of Addictive Behaviors*, 21(4), 534–544. https://doi.org/10.1037/0893-164x.21.4.534

Trelohan, M. (2021). Do Women Engage in Pro-environmental Behaviours in the Public Sphere Due to Social Expectations? The Effects of Social Norm-Based Persuasive Messages. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 33(1), 134–148. https://doi.org/10.1007/s11266-020-00303-9

US Blood Supply Facts. (2018). *American Red Cross*. https://www.redcrossblood.org/donate-blood/how-to-donate/how-blood-donations-help/blood-needs-blood-supply.html#:%7E:text=Each%20year%2C%20an%20estimated%206.8,the%20U.S.%20in%20a%20year.

Who can give blood. (n.d.). *World Health Organization*. https://www.who.int/campaigns/world-blood-donor-day/2020/who-can-give-blood

WHO calls for increase in voluntary blood donors to save millions of lives. (2015, June 11). *World Health Organization*. https://www.who.int/news/item/10-06-2015-who-calls-for-increase-in-voluntary-blood-donors-to-save-millions-of-lives

World Health Organization & International Federation of Red Cross and Red Crescent Societies. (2010). Towards 100% voluntary blood donation: a global framework for action. *World Health Organizatio*n. https://apps.who.int/iris/handle/10665/44359

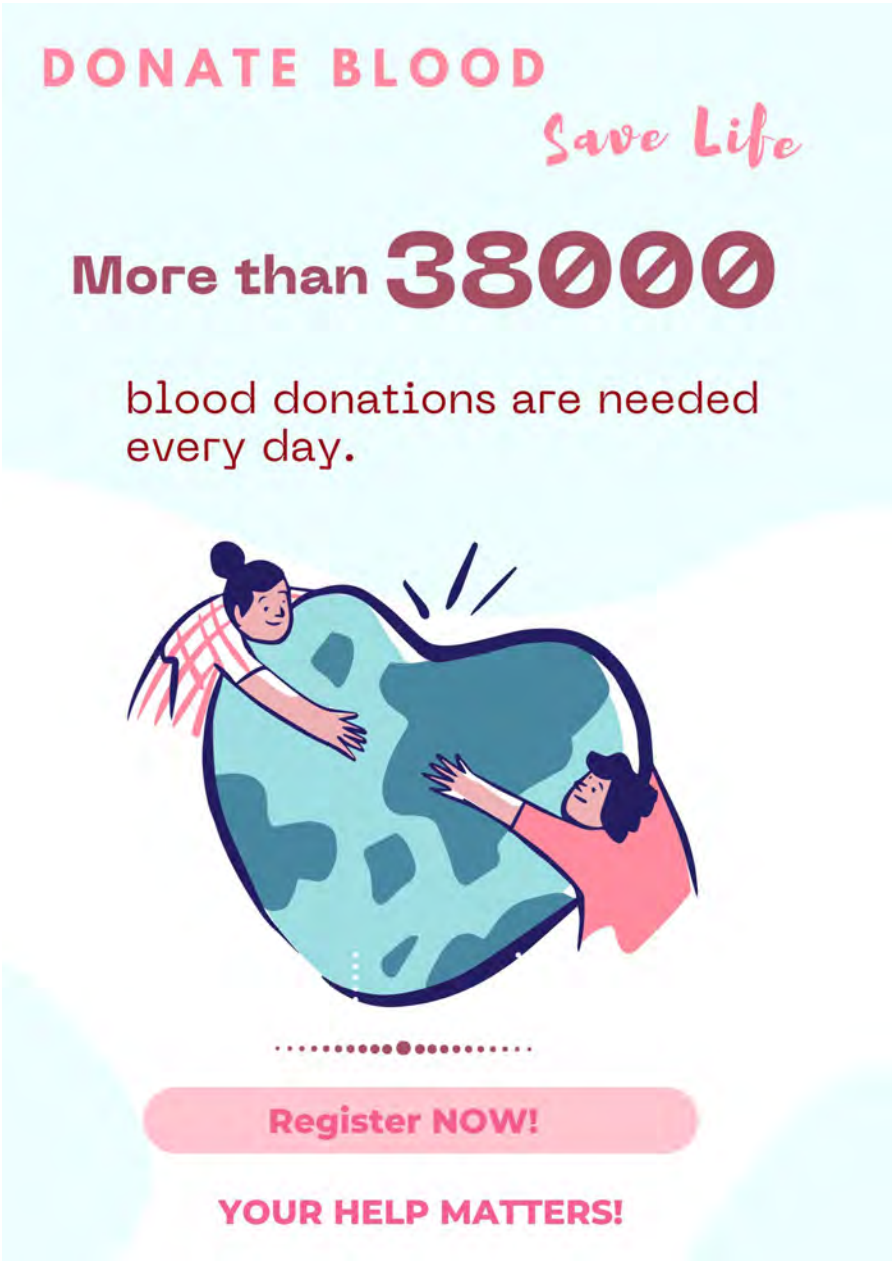
XIE, K. J., MA, J. T., HE, Q., & JIANG, C. M. (2019). Descriptive norms promote willingness to voluntarily donate blood rather than actual blood donation. *Advances in Psychological Science*, 27(6), 1019. https://doi.org/10.3724/sp.j.1042.2019.01019

Yamin, Fei, Lahlou, & Levy. (2019). Using Social Norms to Change Behavior and Increase Sustainability in the Real World: A Systematic Review of the Literature. *Sustainability*, 11(20), 5847. https://doi.org/10.3390/su11205847

Zaller, N., Nelson, K. E., Ness, P., Wen, G., Bai, X., & Shan, H. (2005). Knowledge, attitude and practice survey regarding blood donation in a Northwestern Chinese city. *Transfusion Medicine*, 15(4), 277–286. https://doi.org/10.1111/j.0958-7578.2005.00589.x

Appendix A: Injunctive Norm Advertisement

Appendix B: Control Advertisement



Nature Protecting Nature: The Use of Plant-derived Organic Compounds to Produce Superior Sunscreens that are Both Non-toxic to Coral and Present Reduced Health Risks to Humans

Alexis N. Lindenfelser

Author Background: *Alexis Lindenfelser grew up in the United States and currently attends St. Margaret's Episcopal School in San Juan Capistrano, California in the United States. Her Pioneer research concentration was in the field of chemistry and titled "Solving Materials Science Problems Using Chemistry."*

Abstract

Coral reefs, sometimes called the ocean's rainforests, are one of the most crucial ecosystems on Earth and support a significant amount of aquatic biodiversity. Coral reefs are also popular ecotourism sites and represent economically important regions in island states and territories like Hawaii, the US Virgin Islands, the Philippines, and Indonesia, in addition to protecting some coastal areas from flooding and erosion. Unfortunately, the coral that inhabit these regions are also very fragile organisms. Considering 6,000 tons of sunscreen chemicals wash into the ocean's reefs annually via runoff, wastewater effluent or people entering the water, they have been spotlighted by researchers as a potential culprit for decreased coral health worldwide. These sunscreen chemicals, particularly oxybenzone, octinoxate, avobenzone, octocrylene, octisalate, and homosalate, are organic compounds used in sunscreen to protect human skin from harmful solar rays. However, experiments have shown that, in some cases, the chemical changes undergone by the compounds during UV radiation turns them into phototoxins that disrupt coral health. Additionally, these phototoxins, as well as the atomic makeup of the compounds themselves, might also negatively impact human health. For instance, some studies have shown the potential for the compounds to act as endocrine disruptors in humans.

The negative impact on the environment and potential toxicity towards humans makes it imperative to formulate new sunscreen mixtures with safe and viable ingredients by the discovery, synthesis, or modification of unutilized organic compounds. This review evaluates the potential of a series of plant-

derived and nature-inspired compounds including ubiquinone, quercetin, sinapic acid esters, and thiobarbituric acid derivatives and takes into consideration factors like LogP, antioxidant activity, absorbance, commercial availability, pigmentation, photostability, and toxicity. Overall, most of the introduced compounds, while absorbing in the UVA/UVB range, are best used as photostabilizers or supplements for current compounds which improve the safety and efficacy of current chemical sunscreens until full-spectrum filter replacements can be developed.

1. Introduction

1.1. Environmental Considerations and Sunscreen Overview

Coral reefs are some of Earth's most important ecosystems, both economically and environmentally. Despite covering less than 1% of the ocean floor, coral reefs provide habitats for nearly 1 million different marine species, earning them the distinction as biodiversity hotspots (NPS, n.d.). However, most of the world's coral reefs are endangered, due to both global and local anthropogenic effects. The organisms that build the foundation for coral reefs are very fragile, and their fragility is exacerbated by global warming. The coral polyps (translucent, tiny animals from the Cnidaria family) build a calcium carbonate skeleton to host zooxanthellae which are symbiotic algae that both perform photosynthesis to feed the coral, and give coral its characteristic vibrant colors (National Park Service, n.d.). The symbiotic relationship is mutually benefitting for the coral (which receive a food source) and the zooxanthellae (which receive a place to live). When stressed by temperature, pollution, or other environmental variations, coral expel their zooxanthellae in a process called coral bleaching (NOAA 2021a).This process results in especially vulnerable and weak coral, and corals usually die shortly after bleaching unless they can regain their zooxanthellae (NOAA 2021b; Schneider and Lim 2019). Furthermore, the zooxanthellae of coral often are the ones to sequester toxic sunscreen chemicals, so bleached corals lack this natural defense and are even more susceptible to the phototoxic effects of the UV filters (Vuckovic et al. 2022).

$$CO_2 + H_2O \leftrightarrow H_2CO_3$$
$$(conversion\ of\ CO_2\ into\ carbonic\ acid,\ H_2CO_3)$$

$$H_2CO_3 + H_2O \leftrightarrow HCO_3^- + H_3O^+$$
$$(further\ dissolution\ of\ carbonic\ acid\ into\ bicarbonate\ ions)$$

$$CaCO_3 + 2H_3O^+ \leftrightarrow Ca^{+2} + 2HCO_3^-$$
$$(calcium\ carbonate\ dissolves\ via\ a\ reaction\ with\ hydrodnium\ ions\ to\ become\ calcium\ and\ bicarbonate\ ions)$$

Figure 1. *Chemical reaction pathways by which anthropogenic carbon dioxide in seawater dissolves limestone.*

Additionally, ocean acidification, as a result of anthropogenic carbon dioxide dissolving into carbonic acid in the ocean, dissolves the calcium carbonate (limestone) skeletons that corals build to house the polyps (Osterloff, n.d.). These global changes have been negatively impacting coral for decades, with the planet having lost 50% of its coral reefs since the 1950s (Eddy et al. 2021). In recent years, local anthropogenic impacts have come to international attention. In particular, the culprits responsible for decreased coral health are the chemicals used as UV filters in sunscreens and other cosmetics, which enter ocean ecosystems by runoff from land, such as sewage discharges or garbage leachate, or swimmers entering the water (Fagervold et al. 2019; Downs et al. 2021). This is a problem because of increased sunscreen application, and the increasing popularity of ecotourism (Jordan 2022). It was found that up to 6,000 tons of sunscreen enter reefs annually, concentrated in popular tourist sites and putting those reefs at even greater risk (NPS, n.d.; Jordan 2022). Furthermore, it is nearly impossible for wastewater treatment plants to remove organic UV filters, owing to their high lipophilicity and high organic carbon-water coefficient (Schneider and Lim 2019; Blüthgen et al. 2014).

It is important to first distinguish between chemical sunscreens (below) and physical or mineral-based sunscreens, which use zinc oxide or titanium dioxide particles to dissipate UV rays via reflection (Kimbrough 1997). Nanoparticles are sometimes used in mineral-based sunscreens because they are more desirable for human consumers, however, they are potentially toxic to aquatic life. Studies have shown that the nanoparticles can react with UV light in salt water to form hydrogen peroxide, which causes oxidative stress in phytoplankton (McMahon 2021; Faco et al. 2022; Sánchez-Quiles and Tovar-Sánchez 2014). But, despite mineral-based sunscreens being proven to be safer overall for both coral and humans, physical sunscreens are severely under-adopted by consumers. Their white, pasty, and opaque nature makes them into less convenient, less comfortable, and less desirable sunscreens as compared to chemical sunscreens (Shaath 2010).

Chemical sunscreens contain organic compounds (called UV filters) that can dissipate UV radiation to protect human skin from premature aging, sunburn, and, in the long run, skin cancer (Kimbrough 1997; Environmental Working Group 2022b). They have been the main target of investigations because of their phototoxicity in coral and potential toxicity towards humans.

1.2. General Sunscreen Chemistry

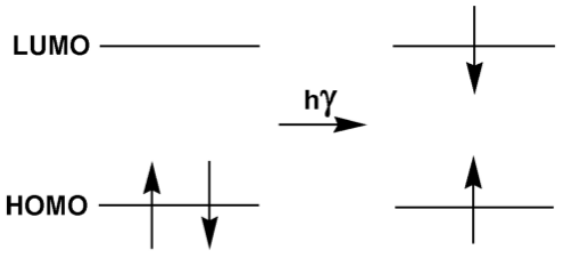


Figure 2. *Promotion of electrons from HOMO to LUMO during absorption of UV photons in organic UV filters (Faco et al. 2022; Shaath 2010).*

When these compounds absorb a UV photon, their electrons are excited, moving them into higher energy orbitals (Faco et al. 2022; Kimbrough 1997; Shaath 2010). As shown in Figure 2, electrons are promoted from the HOMO (highest occupied molecular orbital) to the LUMO (lowest unoccupied molecular orbital). In organic UV filters, the molecule's electrons are returned to the ground state in a series of vibrational transitions that dissipate the absorbed energy, often via photoisomerization (Faco et al. 2022; Kimbrough 1997; Shaath 2010). This process is taken care of by the chromophore of the molecule, which consists of a highly conjugated π-electron system (Faco et al. 2022; Kimbrough 1997).

Thanks to pressure from mounting evidence regarding the dangers of these compounds, as well as pressure from advocacy organizations like the Environmental Working Group, the human-safe concentrations of these compounds are under review by the Food and Drug Administration (Downs et al. 2021; Environmental Working Group 2022b; 2022a; Michele 2021). For example, in the interest of protecting coral health, the U.S. Virgin Islands has banned the sale of sunscreens containing oxybenzone, octocrylene, and octinoxate since March 2020 (McMahon 2021). Aruba has banned sunscreens containing oxybenzone, and Bonaire, Palau, and some ecotourism reserves in Mexico have also instituted policies against the sale of sunscreens with toxic compounds, encouraging tourists to bring only reef-safe or reef-friendly sunscreens (Jordan 2022; McMahon 2021). Hawaii banned the sale of over-the-counter sunscreens and cosmetics containing oxybenzone and octinoxate at the beginning of 2021 (State of Hawaii 2018). This is due in part to studies like those conducted by Mitchelmore et al., which found concentrations of UV filters homosalate, octisalate, oxybenzone and octocrylene in coral tissue, surface water, and sediments in coral reefs around Oahu (Mitchelmore et al. 2019). The concentrations of the compounds were found to be higher in bays with more recreational usage (Mitchelmore et al. 2019).

2. Chemistry of Current UV Filters

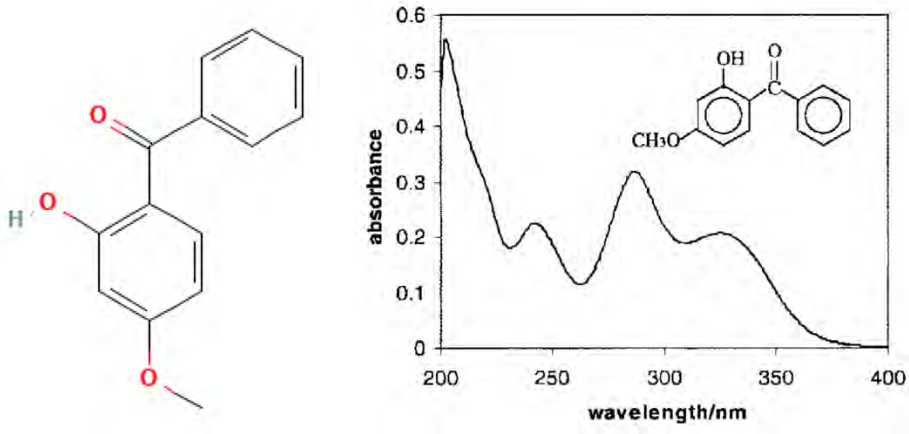
This section will explore six of the organic compounds most commonly found in sunscreens approved in the United States that are currently being investigated for their potential toxicity, and are under the most scrutiny for their potential impact on corals (Michele 2021; Matta et al. 2020; Environmental Working Group 2022b). A summary table of the current UV filters can be found in the Supplementary Materials Section, Table 1.

First, some background on the categorization of UV radiation, which is important in classifying UV filters. UV radiation is divided into three types: UVA, UVB, and UVC. UVA radiation has the longest wavelength, from 315 to 400 nm, making it lower energy compared to the other types, and the least biologically damaging (Kimbrough 1997; WHO 2016). However, sunscreens without ingredients that can dissipate UVA waves have been found to be inadequate, since UVA radiation is capable of triggering oxidative reactions that alter lipids, proteins, and DNA, as well as being immunosuppressive (Kimbrough 1997; Faco et al. 2022). UVB covers wavelengths from 280 to 315 nm and can cause sunburn and trigger carcinogenic reactions (Kimbrough 1997; Faco et al. 2022; WHO 2016). UVC covers from 200 to 280 nm and can kill unicellular organisms upon

irradiation (Kimbrough 1997; Faco et al. 2022; WHO 2016). UVC light is largely blocked from reaching the surface by Earth's ozone layer, so it is more important to focus on finding UVA and UVB filtering chemicals for sunscreens (Kimbrough 1997).

The following sections include discussion of a property known as the octanol-water coefficient, or partition coefficient. It is a way to measure a substance's propensity to dissolve in nonpolar substances, or lipophilicity, by taking the ratio of how much the substance dissolves in octanol (a nonpolar substance) over its water (a polar substance) solubility (Amézqueta et al. 2020). This is important to consider in the discussion of finding new sunscreen ingredients, because sunscreen is commonly used in conjunction with water sports, and sunscreen that dissolves too easily in water would be ineffective. Therefore, it is important to identify compounds with LogP values similar to those of sunscreen chemicals already in circulation.

2.1. Oxybenzone



(**a**) (**b**)

Figure 3. ***(a)*** *Oxybenzone Chemical Structure (NCBI 2022a);* ***(b)*** *Absorbance spectrum for oxybenzone (Salvador et al. 2001).*

Oxybenzone, found in 80% of US non-mineral sunscreens, is a UV filter designed to absorb rays in the UVA/B range, with a maximum wavelength absorbance of 286 nm (Fernandez 2019). Figure 3b shows how wavelengths from 250-380 nm can be absorbed by oxybenzone. Studies have demonstrated that oxybenzone exhibits an excited state intramolecular proton transfer (ESIPT), followed by a molecular rotation that accelerates its return to the ground state after its electrons have been excited by UVA irradiation (Baker et al. 2015; Holt et al. 2020). Essentially, oxybenzone is able to emit the absorbed energy through a series of vibrational transitions, effectively dissipating the absorbed energy and preventing the damaging UV radiation from harming human skin (Kimbrough 1997). Oxybenzone's LogP value is 3.6 (NCBI 2022a).

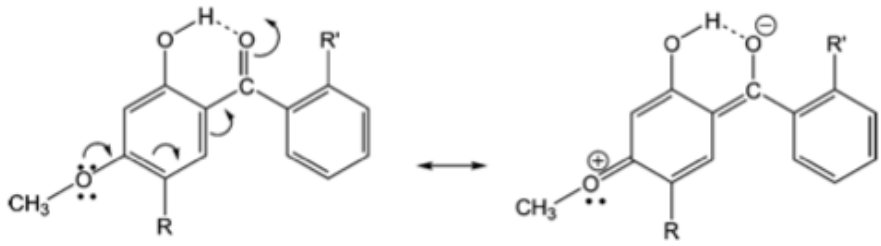


Figure 4. *The electron delocalization in an oxybenzone molecule (Shaath 2010).*

Despite its effectiveness as an organic UV filter, oxybenzone has been flagged in a multitude of studies for its potential toxicity towards both coral and, ironically, the humans that oxybezone is intended to protect. Oxybenzone is also absorbed readily into the skin, having a higher percutaneous absorption than other UV filters, with some studies showing that it was absorbed at a level exceeding the FDA's limit by 516 times (Amézqueta et al. 2020; Matta et al. 2019; 2020). Oxybenzone has been detected in human urine, serum, and breast milk, and it is estimated that 96.8% of the US population is exposed to oxybenzone (Schneider and Lim 2019; S. Q. Wang, Burnett, and Lim 2011; Calafat Antonia M. et al. 2008). Oxybenzone has been shown to be a photoallergen, meaning it (along with other benzophenones) causes allergic skin reactions when exposed to sunlight (Russo et al. 2018). Oxybenzone has also been flagged as a potential endocrine disruptor with the potential to increase the risks of breast cancer and endometriosis, cause decreased testosterone levels in boys with greater exposure, and impact birth weights (Ghazipura et al. 2017; Kunisue et al. 2012; Scinicariello Franco and Buser Melanie C. 2016; Kariagina et al. 2020; Peinado et al. 2021; Rooney et al. 2021; Environmental Working Group 2022b).

In coral, oxybenzone has been shown to have numerous negative impacts ranging from increasing the incidence of coral bleaching by damaging symbiotic zooxanthellae, forming toxic metabolic products during photodegradation, to causing endocrine disruption. Of particular concern is the fact that oxybenzone is also the UV filter found the most frequently and in the highest concentrations in water sources worldwide (Schneider and Lim 2019). In studies conducted on coral planulae, oxybenzone was found to be phototoxic. In other words, its negative impacts are exacerbated by sunlight because UV radiation generates harmful reaction pathways wherein the coral metabolizes oxybenzone into toxic glucosides (Figure 5), mostly through reactions with the phenol group (Downs et al. 2016; Vuckovic et al. 2022). Other studies proved oxybenzone to be a skeletal endocrine disruptor, causing ossification (encasing in one's own skeleton) of coral planula (Downs et al. 2016). Some studies also found that, under certain conditions, oxybenzone could activate the viral lytic cycle in infected zooxanthellae, thus promoting coral bleaching (Danovaro Roberto et al. 2008). The study conducted by Downs et al. in 2016 delved deeper into the potential mechanisms by which oxybenzone-induced damage to zooxanthellae encouraging coral to expel the symbionts. Oxybenzone's genotoxicity, or potential to damage DNA, which is also exacerbated by sunlight, is of special concern since

this can impact reproduction of coral, and thus the survival and re-establishment of corals and potentially other reef organisms (Depledge and Billinghurst 1999; Anderson S L and Wild G C 1994; Downs et al. 2016).

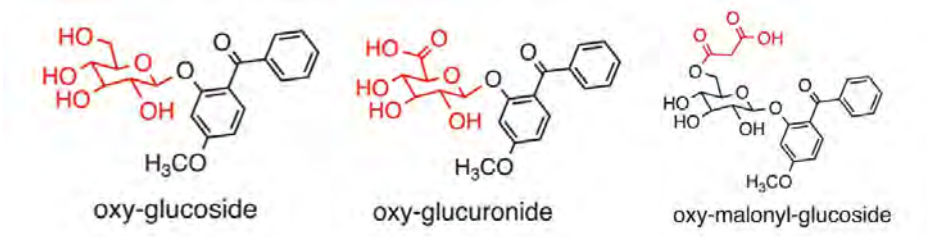


Figure 5. *The metabolization of oxybenzone into harmful glucosides (Vuckovic et al. 2022).*

2.2. Octinoxate

Octinoxate is a commonly used cinnamate UVB (280-315 nm) filter with a maximum absorbance of 311 nm and a LogP of 5.3 (Santos, Miranda, and Esteves Da Silva 2012; NCBI 2022i). To dissipate the energy absorbed from UV photons, octinoxate undergoes a photoisomerization between its cis (Z) and trans (E) isomers (Shaath 2010; Santos, Miranda, and Esteves Da Silva 2012). The E isomer is a more efficient UVB absorber and is more commonly found in sunscreen mixtures (Shaath 2010; Santos, Miranda, and Esteves Da Silva 2012)

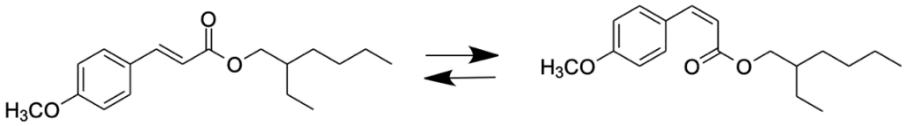


Figure 6. *E and Z isomers of octinoxate (Created using* ChemDraw JS*).*

Octinoxate is classified as an "environmental hazard" by the National Center for Biotechnology Information and is banned in many island countries, as noted in the introduction (NCBI 2022i). Similar to oxybenzone, octinoxate is genotoxic to coral and increases rates of coral bleaching (Smith 2018). Octinoxate's production method is also very fossil-fuel intensive, compounding its negative effects on the environment (Peyrot et al. 2020).

Animal studies have suggested that octinoxate has the potential to disrupt thyroid, androgen, and progesterone endocrine systems (Krause et al. 2012). Again, like oxybenzone, it is a photoallergen. Octinoxate is readily absorbed into human skin, and has been found in blood at levels 16 times FDA maximums (Environmental Working Group 2022b; Matta et al. 2019; 2020).

2.3. Avobenzone

Avobenzone is the most widely implemented UVA (315-400 nm) filter in the world, owing to it having a λ_{max} of 357 nm (Holt et al. 2020; Santos, Miranda, and Esteves Da Silva 2012). Avobenzone's LogP value is 4.8, making it slightly more lipophilic than oxybenzone (NCBI 2022e). When the molecule absorbs UVA radiation, it undergoes a keto-enol tautomerization, and the enol form is energetically favored, attributed to its ability to form intramolecular hydrogen bonds, as shown in Figure 7 (Shaath 2010).

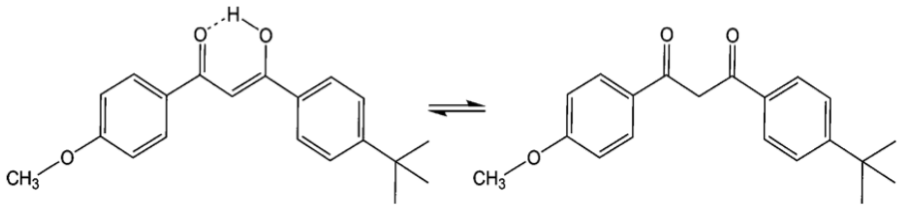


Figure 7. *keto-enol tautomerization of avobenzone (Shaath 2010).*

The main problem with avobenzone is that it is highly photounstable, meaning that avobenzone, particularly the keto form, shown in Figure 9, breaks down into harmful or less-effective UV filtering products after continued exposure to, and absorption of, UV radiation (Afonso et al. 2014; M. S. Díaz-Cruz and Barceló 2009; Giokas, Salvador, and Chisvert 2007; S. M. Díaz-Cruz et al. 2008; La Farré et al. 2008; Richardson et al. 2007; Richardson Susan D. et al. 2010; Hrudey 2009; Santos, Miranda, and Esteves Da Silva 2012). Various dangerous photoproducts of avobenzone fragmentation include substituted benzoic acids, benzils, arylglycoxals, dibenzoylmethanes, and dibenzoylethanes (Huong et al. 2008; Afonso et al. 2014). Because of these photodegradation products, cytotoxic and photoallergic reactions have been associated with avobenzone (Karlsson et al. 2009; Afonso et al. 2014).

Other sunscreen chemicals like octocrylene, homosalate, and octisalate (discussed below) are added to stabilize avobenzone, but may have harmful effects of their own (Lhiaubet-Vallet et al. 2010; Hanson, Gratton, and Bardeen 2006). There is potential to replace octocrylene, homosalate, and octisalate with compounds that act as antioxidants to photostabilize avobenzone, because antioxidants can quench the reactive free radical species generated by avobenzone's keto form, and offer enhanced photoprotection by quenching radical species generated by UV irradiation of the skin (Afonso et al. 2014).

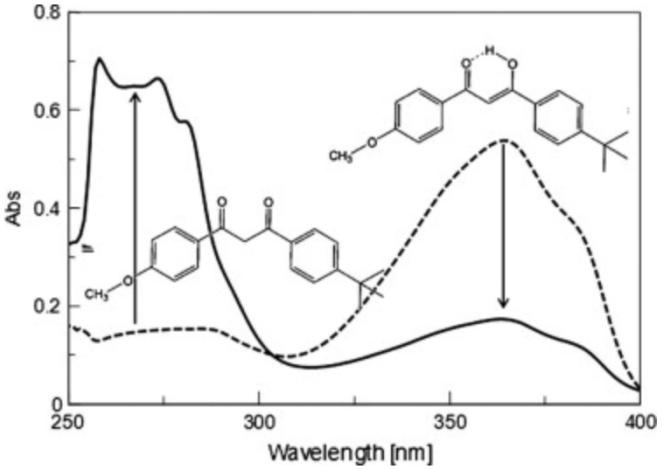


Figure 8. *Avobenzone solution absorption spectrum before (dotted line) and after (solid line) 2h of exposure to UV radiation (Afonso et al. 2014).*

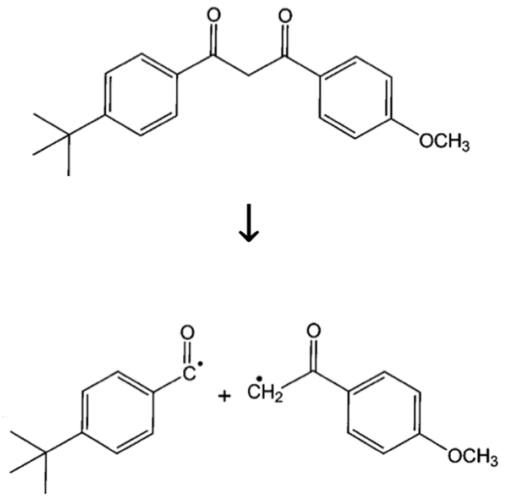


Figure 9. *Degradation of avobenzone's keto form (top) into radicals (Shaath 2010).*

2.4. Octocrylene

Octocrylene is an ester formed by the condensation of 2-ethyhexyl cyanoacetate with benzophenone (Downs et al. 2021; Jing et al. 2018). While octocrylene has some UVB-absorbing capabilities of its own, with a λ_{max} of 303 nm, it is a weak sunscreen on its own. The main reason for its inclusion in sunscreens is to facilitate the stabilization of avobenzone (Santos, Miranda, and Esteves Da Silva 2012; Afonso et al. 2014; Todorov, n.d.). Without octocrylene, avobenzone is degraded by 50% within 1 hour of UV light exposure (Mori and Wang 2021). With

a LogP of 7.1, octocrylene's water resistance properties give it emollient qualities (NCBI 2022d).

However, the continued inclusion of octocrylene in sunscreens is being questioned because of its photoallergenicity, and photoinduced generation of reactive oxygen species (ROS) in human skin cells (Environmental Working Group 2022b; Afonso et al. 2014; Vuckovic et al. 2022). The fact that it is readily absorbed into the skin at rates 14 times the FDA's safety cutoff makes its potential toxicity especially concerning (Hayden et al. 2005; Matta et al. 2020). Studies have also gathered conclusive evidence that octocrylene slowly undergoes a retro-aldol condensation that gives rise to benzophenone as octocrylene-containing products age (Downs et al. 2021). Benzophenone is a known mutagen, carcinogen, and endocrine disruptor, which is banned completely in all personal care products, including sunscreens, under California Proposition 65 (Downs et al. 2021). In addition to benzophenone accumulating as octocrylene-based products age, benzophenone contamination may also result from the production process used to make octocrylene (Downs et al. 2021; Environmental Working Group 2022b).

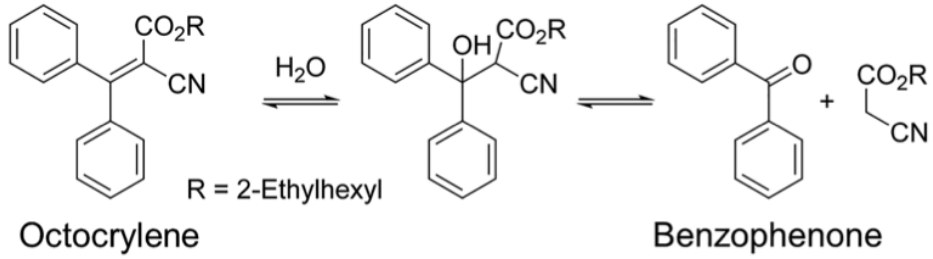


Figure 10. *Retro-aldol condensation of octocrylene that gives rise to benzophenone (Downs et al. 2021).*

As demonstrated in Figure 10 above, the mechanism by which octocrylene degrades to produce benzophenone highlights the importance of new sunscreen compounds not containing two benzene rings linked by a carbonyl. A further reason to replace octocrylene in sunscreens is that the compound has also been found to have significant ecological effects, ranging from endocrine disruption to potential carcinogenicity, and triggering mitochondrial disruption in coral cells (Stien et al. 2018; Blüthgen et al. 2014; Zhang et al. 2016; Gu et al. 2019; Yan et al. 2020; Zdravković et al. 2019). Additionally, octocrylene's association with benzophenone opens up more potential for octocrylene-containing sunscreens to harm coral health and aquatic life (Environmental Working Group 2022b).

2.5. Homosalate and Octisalate

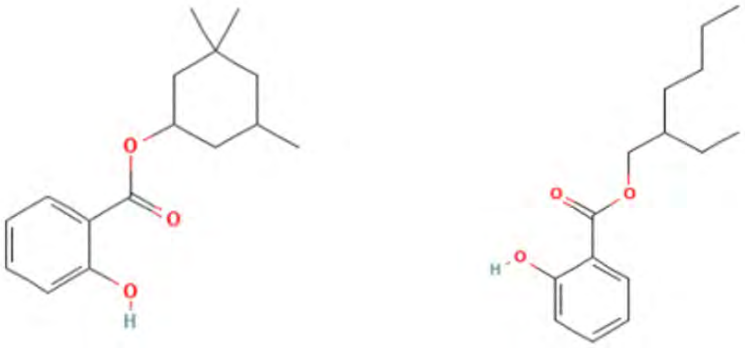


Figure 11. (a) *Homosalate chemical structure, LogP = 5 (NCBI 2022b).*
(b) *Octisalate chemical structure, LogP = 5.7 (NCBI 2022c).*

Homosalate (Figure 11a) is a salicylate molecule absorbing most strongly in the UVB region, with a λ_{max} of 306 nm (Shaath 2010). Owing to the shared phenol group ortho to an ester in their structures, octisalate has similar properties to homosalate, with its λ_{max} being 305 nm and a similar LogP (Shaath 2010). Both assist in the stabilization of avobenzone (Holt et al. 2020). Salicylates (and benzophenones) have the ability to form intramolecular hydrogen bonds, so they, like oxybenzone, undergo an excited state intermolecular proton transfer (ESIPT), specifically keto-enol tautomerization, to return to the ground state after UV irradiation (Holt et al. 2020).

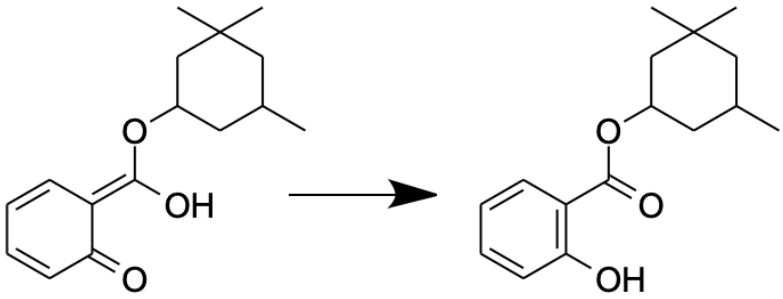


Figure 12. *Proposed keto-enol tautomerization of a homosalate molecule (Created using* ChemDraw JS*).*

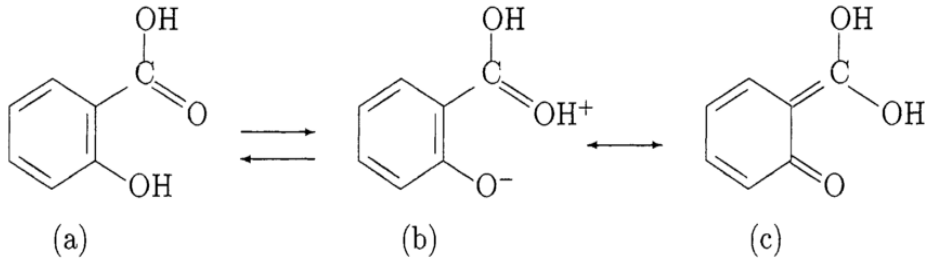


Figure 13. *Primary forms of salicylic acid. This figure inspired the proposed keto-enol tautomerization shown in* Figure 12. *In this diagram, the enol form (a) is likely the only stable form in the ground state, which might also apply to the homosalate molecule (Joshi, Gooijer, and Zwan 2002).*

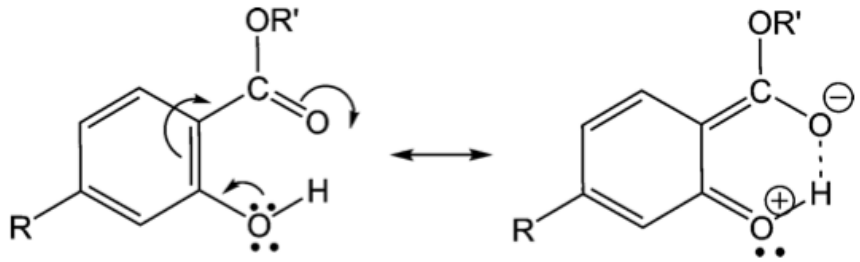


Figure 14. *Electron delocalization in a salicylic molecule (Shaath 2010).*

Both homosalate and octisalate readily cross the skin barrier, and were found in human plasma in levels exceeding FDA safety cutoffs (Matta et al. 2020). Homosalate has additionally been noted as a potential endocrine disruptor, with the ability to disrupt estrogen, androgen, and progesterone levels (Krause et al. 2012). Especially because of its ability to bioaccumulate, homosalate is likely to cause hormonal dysfunction in aquatic life (Yazar and Ertekin 2020). Research has proven that homosalate, and salicylates in general, have cytotoxic and genotoxic effects on MCF-7 cells. Specifically, impacting cell viability, inducing apoptosis, and decreasing glucose consumption in the cells (Yazar and Ertekin 2020; Spitz et al. 2009). Both compounds should be avoided in “reef-safe” sunscreens.

3. Proposed Compounds

With the well-established toxicity of current compounds used in sunscreens towards both humans and coral, the need to propose and explore new organic compounds becomes necessary. As can be extrapolated from previous research, organic compounds with substituted aromatic rings, carbonyls, and carbon-carbon double bonded structures are ideal to make the highly conjugated π-electron system capable of absorbing and dissipating UV radiation. It is also crucial to eliminate benzophenone groups, which are likely to become harmful

photoproducts after the larger compound degrades, as evidenced by octocrylene's fragmentation mechanism. Oxybenzone's reaction to form harmful glucosides indicates that a single-bonded oxygen attached to a benzene ring is also a problematic structure. It has been noted that any avobenzone replacements or proposed ingredients should avoid having a 1,3-diketone group, which break down to yield photoallergenic glyoxals (Professor M. Welker, personal communication, 24 August 2022). It would also be ideal to avoid structures that interact with hormone receptors, but, considering the diverse range of structures exhibited by endocrine disruptors, figuring out a specific functional group to avoid is beyond the scope of this paper.

Finally, there are, of course, more of nature's hidden UV-filtering compounds to explore. For instance, a compound called gadusol, an SPF-providing secretion of fish, which is currently in the early stages of research at Gadusol Laboratories, acquired by Arcaea, is being developed for its potential as a safe and natural sunscreen (Jacoby 2015). This biotech research achieved a breakthrough in August of 2022 after devising a way to produce the compound from genetically modified microbes (Kart 2022). However, with its LogP of -2 (extremely water soluble) and unknown human safety profile, more testing and modification should be done to make it feasible as a sunblock for ocean use (NCBI 2022f). Aside from gadusol, a summary of the proposed compounds presented below is given in the Appendix A, Table A2.

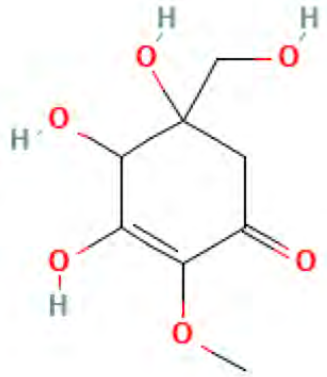


Figure 15. *Gadusol chemical structure (NCBI 2022f).*

3.1. Ubiquinone Proposal

As covered above, avobenzone is photounstable, and ingredients such as homosalate, octisalate, and octocrylene must be added to sunscreen to stabilize avobenzone. These compounds present concerns of their own in terms of human and environmental toxicity. As such, this proposal involves replacing the more dangerous stabilizers with a safe, plant-derived antioxidant that does not degrade or interact with avobenzone (Afonso et al. 2014).

This proposal advocates for the use of ubiquinone (also referred to as

Coenzyme Q10, CoQ10). Ubiquinone is a compound synthesized by most eukaryotic cells and contained in the mitochondria to aid in ATP production, protection of the mitochondrial membrane, and prevention of oxidative damage to DNA (Afonso et al. 2014; Abdul-Rasheed and Farid 2009; Vaghari et al. 2016). With a LogP of 19.4, ubiquinone is extremely insoluble in water, most likely attributed to its extremely long hydrocarbon chain (NCBI 2022h). This characteristic is ideal so that sunscreens do not dissolve in water, however, since the LogP of ubiquinone greatly exceeds those of currently used sunscreen ingredients, further testing would need to be conducted to verify its LogP value does not give rise to solubility problems (e.g., too easily permeates into human skin).

Ubiquinone is commercially available. It is readily obtained via chemical synthesis or biological extraction from plants including soybeans, peanuts, palm oil (plants with the highest concentrations) or animal tissues like beef, pork hearts, and fish, or microbial fermentation. (Vaghari et al. 2016). Biological extraction mechanisms (including microbial fermentation) are both a low cost and environmentally benign options (Vaghari et al. 2016). Genetic modification of microbes could potentially result in higher ubiquinone yield, greater ease of production and lower costs, if future studies are conducted (Vaghari et al. 2016).

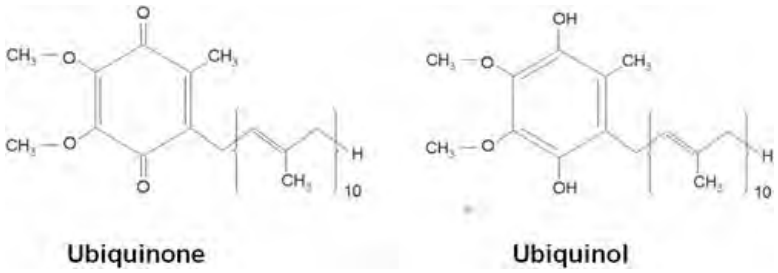


Figure 16. *Chemical structure of ubiquinone (Desbats et al. 2015).*

Ubiquinone acts an antioxidant by accepting the free electrons from free radicals, rendering the radicals harmless as ubiquinone transforms into to its reduced form, ubiquinol (Abdul-Rasheed and Farid 2009). Since avobenzone's keto form, created through photodegradation, results in the production of singlet oxygen species, the use of antioxidants is "a logical and reasonable" strategy to stabilize avobenzone (Afonso et al. 2014). It is worth noting that ubiquinone (the oxidized form) is yellow.

This raises concerns for its potential unaesthetic tinting of sunscreens (and staining of skin). Nonetheless, when reduced, as it will be after quenching free radical species, ubiquinone turns into ubiquinol (reduced form), which is clear (Y. Wang and Hekimi 2020). On the other hand, the compound does readily change between its oxidized and reduced forms, so it cannot be assumed that the color will be completely clear. However, even if it ends up being orange, ubiquinone will be used in sunscreens in such low concentrations—the study conducted by Afonso (Afonso et al. 2014), used a maximum concentration of ubiquinone at 0.01 mg/mL—that it is unlikely to have an impact on the overall color of the sunscreen when combined with majority white and clear ingredients. Further testing would need to be conducted to ensure the color of ubiquinone-

containing sunscreens remains acceptable to consumers.

The study conducted by Afonso (Afonso et al. 2014) also evaluated the potential photostabilization ability of Vitamin C and E (also natural antioxidants), and found ubiquinone to be the most effective photostabilizer. Photostability increases were calculated by taking the area under the curve (AUC) ratio of the UVA range before and after irradiation. At a concentration of 2.5 µg/mL of ubiquinone in solution, there was a 16.7% increase in photostability compared to the control (no antioxidants). However, the relationship between concentration of ubiquinone and photostability increases was not linear, indicating a need to conduct further studies to figure out why this is, and to find the most optimal concentration of ubiquinone.

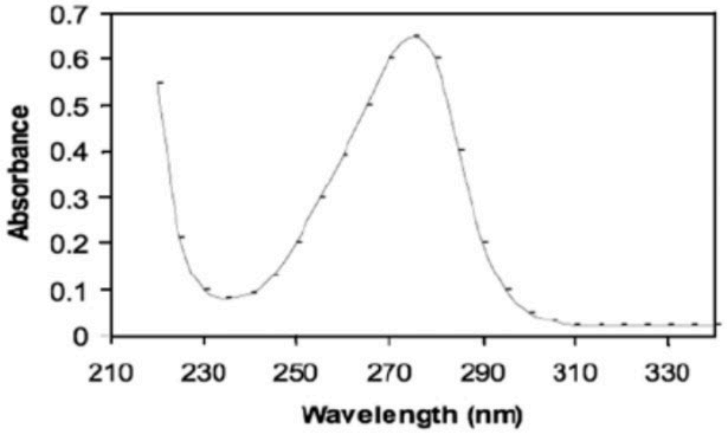


Figure 17. *The absorbance spectrum of ubiquinone dissolved in ethanol, λ_{max} = 275nm (Abdul-Rasheed and Farid 2009).*

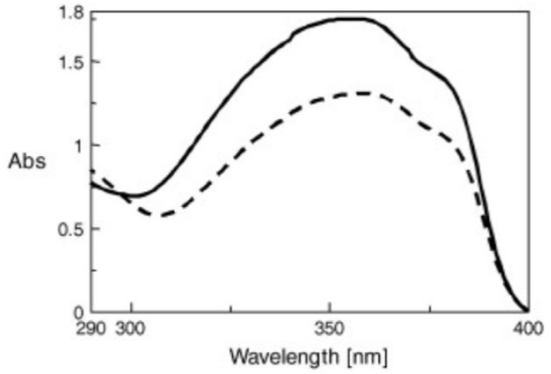


Figure 18. *UV spectra of avobenzone and ubiquinone experimental sunscreen formulation before (solid line) and after (dotted line) irradiation (Afonso et al. 2014).*

While the primary purpose of adding ubiquinone is not UV filtering, the spectrum covering the 250-290 nm range indicates potential for it to absorb a bit of high energy UVB and low energy UVC radiation (Wu et al. 2020). This could have a potential synergistic effect on sunscreen photoprotection, similar to octocrylene usage (Abdul-Rasheed and Farid 2009). This absorbance range does indicate, however, that ubiquinone is photosensitive (responds to sunlight) and suggests the necessity of further testing. Notably, a study conducted by Tournas (Tournas et al. 2006) which tested the photoprotection of ubiquinone used as a cream on its own, did not report any photodegradation products for ubiquinone, but more research is needed.

The curve in Figure 18 can be contrasted to Figure 8 (from the same study), showing a considerably greater decrease in absorbance without ubiquinone. *Note: Figure 8 shows avobenzone absorbance in DMSO, whereas Figure 18 shows absorbance of a sunscreen formulation, but the result, that ubiquinone attenuates decreases in absorbance, is still valid.*

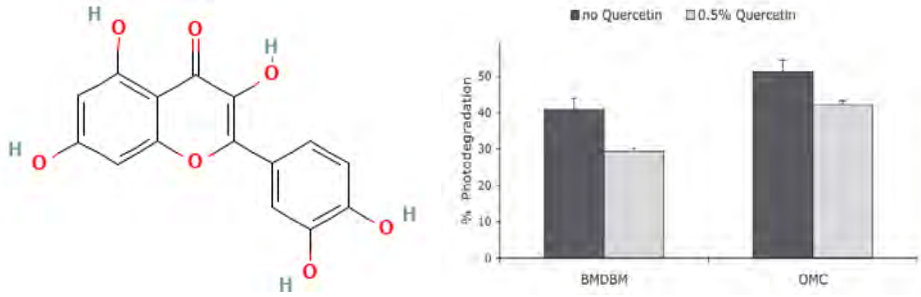
One further note: ubiquinone is extremely unlikely to present toxicity concerns to coral because it is synthesized by the eukaryotic cells of coral themselves and does not contain structures that could present issues upon degradation. However, ubiquinone should be tested in sunscreen formulations that contain more than just avobenzone (as in Afonso et al.) to ensure there are no inadvertent harmful reactions between ubiquinone and other UV filters that hamper the improvement facilitated by the avobenzone-ubiquinone relationship.

3.2. Quercetin Proposal

Other studies have also evaluated the positive impact of similar biologically extracted antioxidants on sunscreen formulations. A study conducted by Scalia (Scalia and Mezzena 2010) investigated sunscreen formulations containing octinoxate and avobenzone with the addition of quercetin, a yellow tinted flavonoid. In this case, quercetin is able to act as triplet-quencher to increase the photostability of avobenzone and octinoxate, while at the same time enhancing photoprotection by deactivating radicals produced by UV radiation (Gaspar and Maia Campos 2006; Herzog, Wehrle, and Quass 2009; Bonda 2005; Scalia and Mezzena 2010).

$$^{3}\text{BM-DBM(K)} + \text{Filter} \longrightarrow \text{BM-DBM(K)} + {}^{3}\text{Filter}$$

Figure 19. *Potential scheme for triplet quenching of avobenzone (BM-DBM) produced radicals by antioxidant filters (Lhiaubet-Vallet et al. 2010).*



(**a**) (**b**)

Figure 20. (a) *Quercetin chemical structure, LogP =1.5 (NCBI 2022g);* **(b)** *The positive impact on photostability by quercetin is conserved over a 3-month storage period (Scalia and Mezzena 2010).*

The incorporation of quercetin also diminished the sunscreen photodecomposition during short-term solar radiation. The stabilization efficacy of quercetin was maximal at 0.5% wt/wt, demonstrating that quercetin's photostabilization-capabilities are concentration dependent. This experimental concentration also mitigates the need to be overly concerned with quercetin's yellow pigmentation because it would be incorporated in concentrations so low that quercetin-containing sunscreens would still be aesthetically acceptable (similar to the logic applied to ubiquinone's pigmentation). To the suggestion of Pioneer scholar Jumana Elnashai, with the high popularity of tinted SPF cosmetic products, there is potential for these pigmented compounds to be applied to that area.

Overall, quercetin was found to be a better stabilizer than octocrylene, even in lower concentrations, and the sunscreen formulations with quercetin still fulfilled official requirements for broadband UVA and UVB protection. It absorbs most strongly in the band from 240-280 nm and 340-440 nm (Buchweitz et al. 2016). *Note: the study was conducted in 2010, so these requirements may have changed since then. Further studies should build on this work to make sure quercetin-avobenzone-octinoxate sunscreens can still meet current requirements.*

The most recent study on quercetin as a photostabilizer was conducted by Gonçalves (Gonçalves et al. 2019) in 2019. The report explored the effect of alkylating quercetin, to produce quercetin 3, 7, 3', 7', 4'-tetraethyl ether, for better sunscreen properties like decreased epidermis penetration and increased lipophilicity. Future experiments can build on the results of these two papers to prepare the best quercetin-based photostabilizer.

While quercetin cannot replace avobenzone and octinoxate as a UV filter, its abilities as a photostabilizing agent can help improve the environmental, efficacy and human safety of current sunscreens until avobenzone and octinoxate filters can be adequately replaced.

Finally, the study by Scalia (Scalia and Mezzena 2010) demonstrates that testing stabilizers in octinoxate and avobenzone mixtures, not just avobenzone

mixed with the proposed stabilizers, is important because octinoxate and avobenzone together have deleterious effects on each other's photostability.

3.3. Sinapate and Sinapic Acid Esters

The following compounds are all structurally related to octinoxate, an artificial cinnamate, which has already been proven to be effective at absorbing UV radiation, so exploring other structurally-related plant-derived (natural) compounds is a sensible next step (Horbury et al. 2020). A commonality between sinapic acid esters is that they possess a phenolic aromatic ring substituted with two methoxy groups, and an alkene bonded to an ester, then a myriad of possible R-groups. Being a cinnamate, octinoxate is similar, but its aromatic ring has only one methoxyl substituent attached. This change in structure may be a cause of octinoxate's toxicity. *Note: Cinnamates and sinapates differ only in a change in the orientation and position of methoxy substituents on the aromatic rings (Dean et al. 2014).*

It is noted that most sinapate esters absorb strongly in the UVA and UVB regions, as they act as "plant sunscreens" (Dean et al. 2014). Antioxidant activities are an added benefit presented by sinapic acid esters, because they increase photoprotection by deactivating free radicals generated by UV radiation (Peyrot et al. 2020). Although not stated in the study by Peyrot (Peyrot et al. 2020), the antioxidant properties of these molecules may also assist with the stabilization of compounds like avobenzone, whose photodegradation results in the production of free radicals. They are expected to have long-term photostability and photoprotectivity due to the trans-cis isomerization undergone after irradiation (Horbury et al. 2020).

Many sinapic acid esters can be derived from readily-available sinapoyl malate, a molecule found in the upper epidermis of plant leaves to defend against the potential harms of UV photons reaching the plant, such as oxidative DNA and tissue damage and inhibition of photosynthesis and growth (Dean et al. 2014). Sinapic acids are particularly abundant in a family of plants called Brassica, which includes the likes of cabbages and mustards (Nguyen et al. 2021)

The following proposal presents a few sinapate ester variations for usage in sunscreens but is mostly a stepping-stone for the development of more suitable custom molecules to find the best sinapate-based UV filter.

3.3.1. DHDES Proposal

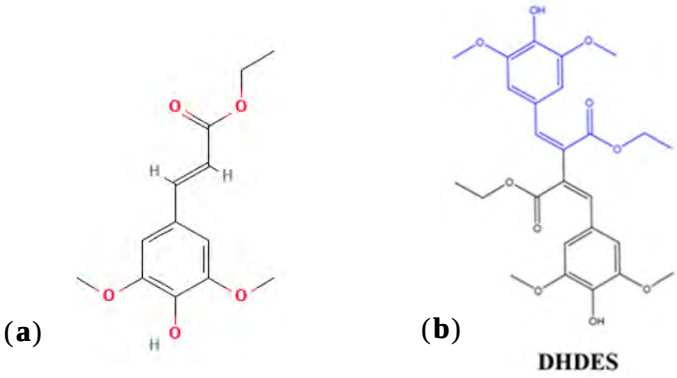


Figure 21.
(a) *Ethyl-sinapatechemical structure (NCBI 2022j);*

(b) *Dehydrodiethylsinapate chemical structure (Horbury et al. 2020)*

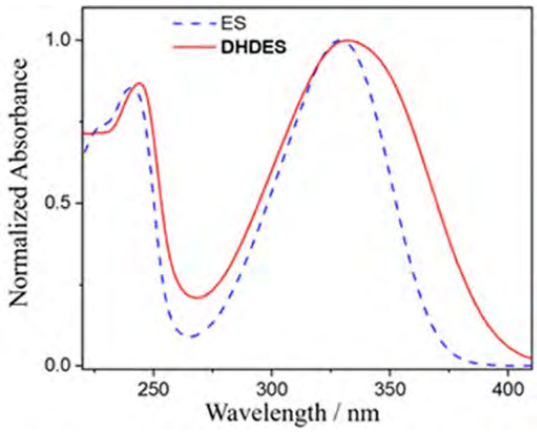


Figure 22. *Absorbance of dehydrodiethylsinapate (DHDES) as compared to ethylsinapate (ES) (Horbury et al. 2020).*

A study conducted by Horbury (Horbury et al. 2020) tested the potential for an ethyl sinapate dimer, so called dehydrodiethylsinapate (DHDES) to be a UV filter with the potential to replace octinoxate, without the harmful genotoxic or endocrine-disrupting qualities. DHDES is formed via the dimerizartion of ethyl sinapate (Figure 21b), in order to increase the degree of π-electron conjugation in the chromophore unit, therefore broadening the range wavelengths that can be absorbed by the molecule (Horbury et al. 2020). A methylated version of DHDES (Me-DHDES) was also tested, and was found to have a similar photostability and UV absorbance range (Horbury et al. 2020)

With the expanded UVA absorbance of DHDES, its photostability was then evaluated using transient electronic absorption spectroscopy in the study (Horbury et al. 2020). Since a small fraction of DHDES was missing post-irradiation, DHDES probably degrades after persistent UV exposure (Horbury et al. 2020). It is expected that quinone methide is the photodegradation product for DHDES, and a tricyclic molecule for a Me-DHDES (Figure 23). The study notes that these photodegradation products will also degrade, giving rise to other unknown molecular species. The toxicity of these photodegradation products would have to be evaluated in further studies.

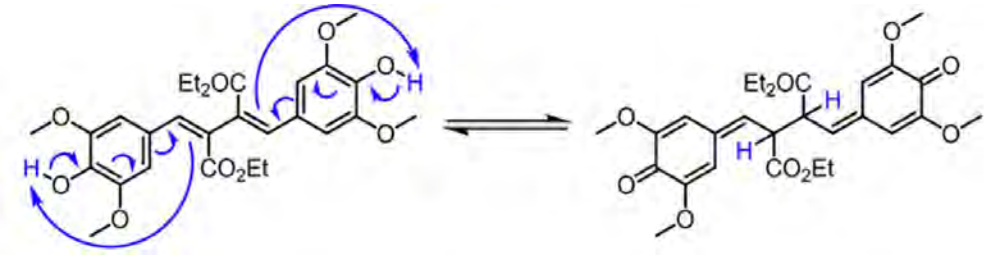


Figure 23. *Proposed photodegradation method of DHDES (Horbury et al. 2020).*

A few other considerations for the DHDES proposal are that it most likely undergoes an ultrafast electronic excited state relaxation via photodimerization between its cis-cis and trans-trans isomers, similar to other sinapate analogues. Secondly, ethyl-sinapate has a LogP value of 2.2, indicating that it is rather soluble (NCBI 2022j). However, the greater number of hydrocarbons in DHDES would probably make it less water soluble than ethyl-sinapate, which is ideal for sunscreens. Ethyl-sinapate is also commercially available, so its dimerization to produce DHDES should not present a significant hurdle.

3.3.2. Other Structurally Related Compounds

In a study conducted by Peyrot et al., sinapic acid ester derivatives were synthesized and compared to octinoxate in an effort to find suitable replacement for the compound that balanced UV filtering ability, continued efficacy after UV absorption with antioxidant activity. They created various experimental compounds via a multitude of detailed reaction pathways. The first group of compounds (octinoxate analogues, compounds 1-3) tested had a similar structure to ethyl-sinapate, but all groups on the substituted ring were methylated (i.e., no hydroxyl), and the alkane chain was lengthened. These compounds had λ_{max} values shifted towards longer wavelengths in the UVA range as compared to octinoxate. Their changes in absorbance abilities after UV exposure as compared to octinoxate was about the same. The next round of compounds tested had modifications made to the R-group connected to the ester (aliphatic sinapate, compounds 4-9). These generally had better absorbency as compared to the first round of compounds, and their absorbance was also shifted to longer wavelengths (UVA) as compared to octinoxate. Loss of absorbance varied. The third group of compounds (phenolic esters, compounds 10-14) involved the addition of aromatic rings to the ester groups. This group was found to have the strongest absorbance, but also the greatest loss of absorbance after UV exposure, with one compound experiencing an 85% loss of absorbance. The final series (compounds 15-17) featured a ketone and two amide derivatives, which had varied UV spectra, but the amide compounds were found to have greater antioxidant properties.

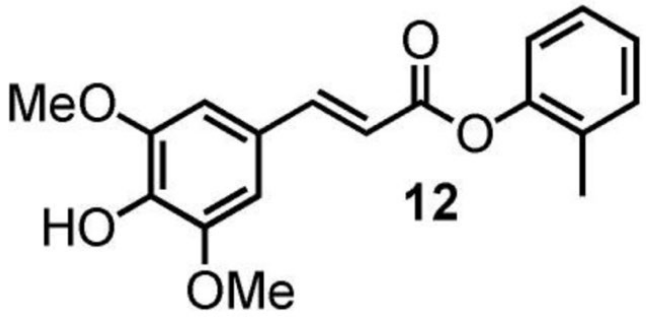


Figure 24. *Compound 12 from the phenolic ester series in the experiment conducted by Peyrot (Peyrot et al. 2020).*

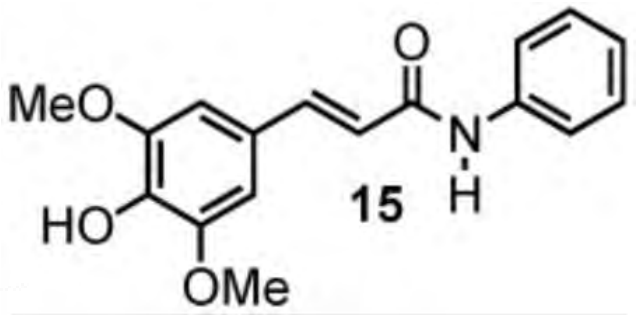


Figure 25. *Compound 15 from the amide/ketone series in the experiment conducted by Peyrot (Peyrot et al. 2020).*

Wavelengths for these compounds are still shifted to cover wavelengths in the UVA range, so they would most likely need to be combined with UVB filters in sunscreen formulations in order to reach full spectrum coverage. Also, the UVA spectrum is expected because the addition of more aromatic groups increases π-electron conjugation, and more conjugation results in lower wavelength absorption in the chromophore. The aliphatic sinapate group demonstrated the importance of having considerable steric hindrance in order to promote cis/trans isomerization and thus better UV absorbing-abilities (Peyrot et al. 2020). However, the phenolic ester group indicates that there is a limit to the amount of steric hindrance that can be introduced before the absorbency actually starts to decrease (Peyrot et al. 2020).

Additionally, it was found that the presence of methoxy substituents (-OMe) on the aromatic ring significantly decreased the interaction of molecules with endocrine receptors, thus lowering the risk of methoxy-substituted molecules to cause endocrine disruption (Hong et al. 2016; Peyrot et al. 2020). As another note, the presence of a phenol is likely to increase water solubility. It is important to consider the impact on LogP made when structural changes are introduced to make sure compounds with more limited water solubility are used.

With an understanding of the property changes caused by certain structural changes in sinapic acid esters, future studies can proceed by combining the most desirable structural features to produce the most ideal compounds. These ideal compounds would balance UV filtering ability with antioxidant properties, reasonable photodegradation, and minimization of structures known to interact with endocrine receptors or increase the risk of producing dangerous decay products (as in the case of the octocrylene-benzophenone degradation mechanism).

3.4. Thiobarbituric Acid Proposal

Using studies conducted by Horbury (Horbury et al. 2020) and Peyrot (Peyrot et al. 2020), which proposed the use of p-hydroxycinnamic acid-based UV filters, a study conducted by Rioux (Rioux et al. 2022) furthered the research by adding thiobarbituric acid. The synthesis method combined the sinapic acid base with a thiobarbituric acid in a high-yielding Knoevenagel condensation. Rioux (Rioux et al. 2022) made adjustments to the synthesis method so that it would be more eco-

friendly and better for human health. This "greener pathway" involved performing the study in water without the presence of a catalyst or toxic chemical solvents, and at room-temperature to conserve energy (Peyrot et al. 2020; Kaupp, Reza Naimi-Jamal, and Schmeyers 2003). The sulfur and nitrogen groups introduced with thiobarbituric acid were expected to shift the UV spectrum into the range of 380-500 nm, which covers a bit of UVA and areas of the blue visible spectrum.

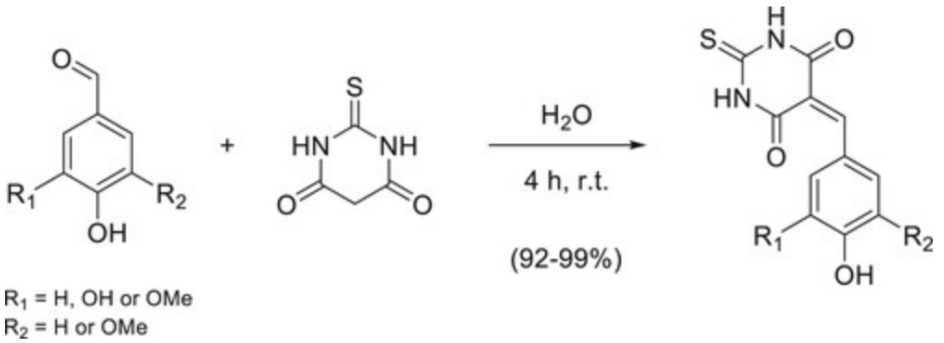


Figure 26. *Knoevenagel condensation of sinapic and thiobarbituric acid (Rioux et al. 2022).*

The λ_{max} values for the thiobarbituric derivates range from 399-445 nm, reaching long wavelengths on the cusp of the UVA region. The paper does report that, however, the derivatives had broadband (from UV to visible region) absorption profiles, and that thiobarbituric derivatives could be a replacement for UV filters like oxybenzone. These absorption profiles indicate that the derivatives are likely pigmented, which is not ideal for sunscreens. If used in low enough concentrations (similar to the mitigation technique for ubiquinone's orange color in Proposal 3.1.), this might not be an issue. Despite these disadvantages, an added feature of the thiobarbituric derivatives is that they expected a much lower loss of absorbance after irradiation as compared to the sinapic acid esters tested from Proposal 3.2., with the loss in absorbance ranging from only 0.1-4.6% after 2 hours, whereas some sinapic acid ester derivatives lost up to 85% of their absorption after one hour (*Note: Irradiation techniques might have been different, so these are not completely comparable results, but the point still stands*).

4. Further Testing

A myriad of *in silico* (computer simulation) techniques could be employed to predict the environmental and biological fate of newly proposed ingredients, which mostly relies on comparison to compounds with similar structures (Rioux et al. 2022). To evaluate potential mutagenicity and carcinogenicity of compounds, the method used by Rioux (Rioux et al. 2022), could be employed, which used the Toxicity Estimation Software Tool (TEST), which is used by the EP, and VEGA platform, then preformed quantitative analysis to re-scale and interpret the value. The potential endocrine-disrupting abilities of compounds

could be measured by using a series of machine learning models from the VEGA platform, which gives a range of predictions (high to low probability), and a reliability for those predictions (high to low reliability) for a molecule to interact with endocrine receptors (Rioux et al. 2022). The strategy in Rioux's study (Rioux et al. 2022) for investigating acute and short-term toxicity could also be of use. Studies by Matta (Matta et al. 2020; 2019) demonstrated a way to determine the relative propensity of molecules to be absorbed into the skin, as a method for discussing potential systemic exposure of people to sunscreen ingredients. The volunteers selected for the study had a diverse range of ages, races, and sex to see if results were consistent among individuals of different groups.

A DPPH (2,2-diphenyl-1-picryhydrayl) assay could be used to determine antiradical properties. The method uses DPPH as a free radical and measures its disappearance over the time in the presence of antiradical compounds, allowing for the calculation of EC50 values (van Schijndel et al. 2017; Glück et al. 2018; Rioux et al. 2022). As a standard procedure, UV absorbance spectra and photodegradation (measured by loss of absorbance or missing compound post-chromatography) could be predicted using UV spectroscopy (Rioux et al. 2022; Horbury et al. 2020; Peyrot et al. 2020).

In order to get a sense of the potential environmental risks posed by certain molecules, factors like their bioaccumulation (buildup in organisms), biodegradability (ability to break down naturally) and environment persistence should be considered, using models provided by the VEGA platform (Rioux et al. 2022). As a further note, chemicals that show favorable results when evaluated for their human safety are likely to also be safe for coral, based on the fact that the currently used compounds are bad for coral and have negative impacts on humans.

Most importantly, a study of the impact of the proposed compounds on aquatic life must be conducted. *In vivo* studies could be modeled after those in the paper by Vuckovic (Vuckovic et al. 2022) which evaluated oxybenzone impacts on *Aiptasia* (a symbiotic anemone) and mushroom coral *Discosoma* under varying lighting conditions and tracked coral survival over time. The methods utilized in a study conducted by Downs (Downs et al. 2021) investigating oxybenzone's effects on coral planula are also suggested for use in generating toxicity profiles for new compounds.

When applicable, the tests should be repeated both with the proposed ingredient on its own, and with the ingredient in an experimental sunscreen formulation, to monitor for potential inadvertent side reactions and interactions between ingredients that might change the properties. Overall, it is suggested that, firstly, a sunscreen formulation of avobenzone and octinoxate with ubiquinone as a photostabilizer is first tested for pigmentation and compared to traditional avobenzone sunscreens. In a second round of testing, DHDES, created via the methods proposed in Horbury's study (Horbury et al. 2020) could be incorporated for added UV-filtering. The absorbency should be tested to see if concentrations of octinoxate and avobenzone can be decreased, their UV filtering-abilities made up for by the presence of DHDES.

5. Conclusion

This review presents a myriad of solutions to the problem of toxic UV filters being used in sunscreens; a problem that has been exemplified in recent years because of the proven deleterious impacts to coral. Coral's added vulnerability due to ocean warming and ocean acidification, as well as their extreme importance in harboring ocean biodiversity, further exacerbates the need to try as many feasible methods as possible to preserve coral reefs, including sunscreen reformulation plans. Gathering as much information on the proven toxicity mechanisms of current UV filters is critical to understanding what is needed, and what must be avoided, in new UV filters for sunscreens. Testing and evaluating proposed safer UV filters through *in silico*, *in vitro*, and *in vivo* methods, as presented here, are an important step to guide the formulating of sunscreens better for humans in terms of both health safety and photoprotection, and that are, of course, non-toxic to aquatic life and coral reefs. Most likely, a combination of solutions will be necessary, and compounds will probably be used in succession as further testing is done and closer to ideal and multi-purpose combinations are discovered. For example, substituting in safer stabilizers with current formulations might be a preliminary step before current UV filters can be completely replaced with compounds that both have full spectrum coverage, ideal LogP and antioxidant properties, and are non-toxic. In conclusion, nature can protect nature with plant-derived ingredients to provide sun protection for humans while not posing risks to the environment or other species needed in the delicate balance of life on Earth.

References

Abdul-Rasheed, Omar F., and Yahya Y. Farid. 2009. "Development of a New High Performance Liquid Chromatography Method for Measurement of Coenzyme Q10 in Healthy Blood Plasma." *Saudi Medical Journal* 30 (9): 1138–43.

Afonso, S., K. Horita, J.P. Sousa e Silva, I.F. Almeida, M.H. Amaral, P.A. Lobão, P.C. Costa, Margarida S Miranda, Joaquim C. G. Esteves Da Silva, and J.M. Sousa Lobo. 2014. "Photodegradation of Avobenzone: Stabilization Effect of Antioxidants." *Journal of Photochemistry and Photobiology B: Biology* 140 (November): 36–40. https://doi.org/10.1016/j.jphotobiol.2014.07.004.

Amézqueta, S., X. Subirats, C. Ràfols, M. Rosés, and E. Fuget. 2020. "Chapter 6 - Octanol-Water Partition Constant." *Handbooks in Separation Science, Liquid-Phase Extraction*, 183–208. https://doi.org/10.1016/B978-0-12-816911-7.00006-2.

Anderson S L and Wild G C. 1994. "Linking Genotoxic Responses and Reproductive Success in Ecotoxicology." *Environmental Health Perspectives* 102 (suppl 12): 9–12. https://doi.org/10.1289/ ehp.9410 2s129.

Baker, Lewis A., Michael D. Horbury, Simon E. Greenough, Philip M. Coulter, Tolga N. V. Karsili, Gareth M. Roberts, Andrew J. Orr-Ewing, Michael N. R. Ashfold, and Vasilios G. Stavros. 2015. "Probing the Ultrafast Energy Dissipation Mechanism of the Sunscreen Oxybenzone after UVA Irradiation." *The Journal of Physical Chemistry Letters* 6 (8): 1363–68. https://doi.org/10.1021/acs.jpclett.5b00417.

Blüthgen, Nancy, Nicole Meili, Geraldine Chew, Alex Odermatt, and Karl Fent. 2014. "Accumulation and Effects of the UV-Filter Octocrylene in Adult and Embryonic Zebrafish (Danio Rerio)." *Science of The Total Environment* 476–477 (April): 207–17. https://doi.org/10.1016/j.scitotenv.2014.01.015.

Bonda, C. A. 2005. "The Photostabilty of Organic Sunscreen Actives." In *In Sunscreens*, edited by Nadim A. Shaath, 323–45. Boca Raton, Florida: Taylor Francis Group.

Buchweitz, Maria, Paul A. Kroon, Gillian T. Rich, and Peter J. Wilde. 2016. "Quercetin Solubilisation in Bile Salts: A Comparison with Sodium Dodecyl Sulphate." *Food Chemistry* 211 (November): 356–64. https://doi.org/10.1016/j.foodchem.2016.05.034.

Calafat Antonia M., Wong Lee-Yang, Ye Xiaoyun, Reidy John A., and Needham Larry L. 2008. "Concentrations of the Sunscreen Agent Benzophenone-3 in Residents of the United States: National Health and Nutrition Examination Survey 2003–2004." *Environmental Health Perspectives* 116 (7): 893–97. https://doi.org/10.1289/ehp.11269.

Danovaro Roberto, Bongiorni Lucia, Corinaldesi Cinzia, Giovannelli Donato, Damiani Elisabetta, Astolfi Paola, Greci Lucedio, and Pusceddu Antonio. 2008. "Sunscreens Cause Coral Bleaching by Promoting Viral Infections." *Environmental Health Perspectives* 116 (4): 441–47. https://doi.org/10.1289/ehp.10966.

Dean, Jacob C., Ryoji Kusaka, Patrick S. Walsh, Florent Allais, and Timothy S. Zwier. 2014. "Plant Sunscreens in the UV-B: Ultraviolet Spectroscopy of Jet-Cooled Sinapoyl Malate, Sinapic Acid, and Sinapate Ester Derivatives." *Journal of the American Chemical Society* 136 (42): 14780–95. https://doi.org/10.1021/ja5059026.

Depledge, M.H, and Z Billinghurst. 1999. "Ecological Significance of Endocrine Disruption in Marine Invertebrates." *Marine Pollution Bulletin* 39 (1): 32–38. https://doi.org/10.1016/S0025-326X(99)00115-0.

Desbats, Maria Andrea, Giada Lunardi, Mara Doimo, Eva Trevisson, and Leonardo Salviati. 2015. "Genetic Bases and Clinical Manifestations of Coenzyme Q10 (CoQ10) Deficiency." *Journal of Inherited Metabolic Disease* 38 (1): 145–56. https://doi.org/10.1007/s10545-014-9749-9.

Díaz-Cruz, M. Silvia, and Damiá Barceló. 2009. "Chemical Analysis and Ecotoxicological Effects of Organic UV-Absorbing Compounds in Aquatic Ecosystems." *Trends in Analytical Chemistry* 28: 708–17.

Díaz-Cruz, Silvia M., Marta Llorca, Damià Barceló, and Damià Barceló. 2008. "Organic UV Filters and Their Photodegradates, Metabolites and Disinfection by-Products in the Aquatic Environment." *Advanced MS Analysis of Metabolites and Degradation Products - I* 27 (10): 873–87. https://doi.org/10.1016/j.trac.2008.08.012.

Downs, C. A., Joseph C. DiNardo, Didier Stien, Alice M. S. Rodrigues, and Philippe Lebaron. 2021. "Benzophenone Accumulates over Time from the Degradation of Octocrylene in Commercial Sunscreen Products." *Chemical Research in Toxicology* 34 (4): 1046–54. https://doi.org/10.1021/acs.chemrestox.0c00461.

Downs, C A, Esti Kramarsky-Winter, Roee Segal, John Fauth, Sean Knutson, Omri Bronstein, Frederic R Ciner, et al. 2016. "Toxicopathological Effects of the Sunscreen UV Filter, Oxybenzone (Benzophenone-3), on Coral Planulae and Cultured Primary Cells and Its Environmental Contamination in Hawaii and the U.S. Virgin Islands." *Archive of Environmental Contamination Toxicology* 70 (2): 265–88. https://doi.org/10.1007/s00244-015-0227-7.

Eddy, Tyler D., Vicky W. Y. Lam, Gabriel Reygondeau, Andres M. Cisneros-Montemayor, Krista Greer, Maria Lourdes D. Palomares, John F. Bruno, Yoshitaka Ota, and William W.L. Cheung. 2021. "Global Decline in Capacity of Coral Reefs to Provide Ecosystem Services." *OneEarth* 4 (9): 1278–85. https://doi.org/10.1016/j.oneear.2021.08.016.

Environmental Working Group. 2022a. "EWG Sunscreen Guide: EWG's 16th Annual Guide to Sunscreen." https://www.ewg.org/sunscreen/report/the-trouble-with-sunscreen-chemicals/.

———. 2022b. "EWG Sunscreen Guide: The Trouble with Ingredients in Sunscreens." https://www.ewg.org/sunscreen/report/the-trouble-with-sunscreen-chemicals/.

Faco, Hazel Anika L., Maxyl Joshua Guillermo, Gresha Sheine S. Larido, Zyarrah Zaida Pangolima, Deserie Joy Vanzuela, and Erwin M. Faller. 2022. "Potential Systemic Toxicity of UV Filters in Sunscreen: A Review." *International Journal of Research Publication and Reviews* 3 (5): 3176–91.

Fagervold, S. K., A. S. Rodrigues, C. Rohée, R. Roe, M. Bourrain, D. Stien, and P. Lebaron. 2019. "Occurrence and Environmental Distribution of 5 UV Filters During the Summer Season in Different Water Bodies." *Water, Air, & Soil Pollution* 230 (7): 172. https://doi.org/10.1007/s11270-019-4217-7.

Fernandez, John. 2019. "Sunscreen Provides Vital Protection Year-Round against the Sun's Ultraviolet (UV) Rays, Which Are Linked to the Skin's Premature Aging, Sunburn and a Person's Risk of Developing Skin Cancer." *Baptist Health South Florida* (blog). June 11, 2019. https://baptisthealth.net/baptist-health-news/sunscreen-ingredients-clarifying-the-confusion/.

Gaspar, L.R., and P.M.B.G. Maia Campos. 2006. "Evaluation of the Photostability of Different UV Filter Combinations in a Sunscreen." *International Journal of Pharmaceutics* 307 (2): 123–28. https://doi.org/10.1016/j.ijpharm.2005.08.029.

Ghazipura, Marya, Richard McGowan, Alan Arslan, and Tanzib Hossain. 2017. "Exposure to Benzophenone-3 and Reproductive Toxicity: A Systematic Review of Human and Animal Studies." *Reproductive Toxicology* 73 (October): 175–83. https://doi.org/10.1016/j.reprotox.2017.08.015.

Giokas, Dimosthenis L., Amparo Salvador, and Alberto Chisvert. 2007. "UV Filters: From Sunscreens to Human Body and the Environment." *TrAC Trends in Analytical Chemistry* 26 (5): 360–74. https://doi.org/10.1016/j.trac.2007.02.012.

Glück, Josephin, Thorsten Buhrke, Falko Frenzel, Albert Braeuning, and Alfonso Lampen. 2018. "In Silico Genotoxicity and Carcinogenicity Prediction for Food-Relevant Secondary Plant Metabolites." *Food and Chemical Toxicology* 116 (June): 298–306. https://doi.org/10.1016/j.fct.2018.04.024.

Gonçalves, Marlucy da Cruz, Viviane Martins R. dos Santos, Jason Guy Taylor, Fernanda Barçante Perasoli, Orlando David H. dos Santos, Ana Carolina S. Rabelo, Joamyr Victor Rossoni Jr., Daniela Caldeira Costa, and Thiago Cazati. 2019. "Preparation and characterization of a quercetin-tetraethyl ether-based photoprotective nanoemulsion." *New Chemistry* 42 (4): 365–70. https://doi.org/10.21577/0100-4042.20170345.

Gu, Jiayuan, Tao Yuan, Ni Ni, Yuning Ma, Zhemin Shen, Xiaodan Yu, Rong Shi, Ying Tian, Wei Zhou, and Jun Zhang. 2019. "Urinary Concentration of Personal Care Products and Polycystic Ovary Syndrome: A Case-Control Study." *Environmental Research* 168 (January): 48–53. https://doi.org/10.1016/j.envres.2018.09.014.

Hanson, Kerry M., Enrico Gratton, and Christopher J. Bardeen. 2006. "Sunscreen Enhancement of UV-Induced Reactive Oxygen Species in the Skin." *Free Radical Biology and Medicine* 41 (8): 1205–12. https://doi.org/10.1016/j.freeradbiomed.2006.06.011.

Hayden, C.G.J., S.E. Cross, C. Anderson, N.A. Saunders, and M.S. Roberts. 2005. "Sunscreen Penetration of Human Skin and Related Keratinocyte Toxicity after Topical Application." *Skin Pharmacology and Physiology* 18 (4): 170–74. https://doi.org/10.1159/000085861.

Herzog, Bernd, Monika Wehrle, and Katja Quass. 2009. "Photostability of UV Absorber Systems in Sunscreens†." *Photochemistry and Photobiology* 85 (4): 869–78. https://doi.org/10.1111/j.1751-1097.2009.00544.x.

Holt, Emily L., Konstantina M. Krokidi, Matthew A. P. Turner, Piyush Mishra, Timothy S. Zwier, Natércia N. Rodrigues, and Vasilios G. Stavros. 2020. "Insights into the Photoprotection Mechanism of the UV Filter Homosalate." *Physical Chemistry Chemical Physics* 22 (June): 15509–19. https://doi.org/10.1039/D0CP02610G.

Hong, Huixiao, Benjamin G. Harvey, Giuseppe R. Palmese, Joseph F. Stanzione, Hui W. Ng, Sugunadevi Sakkiah, Weida Tong, and Joshua M. Sadler. 2016. "Experimental Data Extraction and in Silico Prediction of the Estrogenic Activity of Renewable Replacements for Bisphenol A." *International Journal of Environmental Research and Public Health* 13 (7). https://doi.org/10.3390/ijerph13070705.

Horbury, Michael D., Matthew A. P. Turner, Jack S. Peters, Matthieu Mention, Amandine L. Flourat, Nicholas D. M. Hine, Florent Allais, and Vasilios G. Stavros. 2020. "Exploring the Photochemistry of an Ethyl Sinapate Dimer: An Attempt Toward a Better Ultraviolet Filter." *Frontiers in Chemistry* 8. https://doi.org/10.3389/fchem.2020.00633.

Hrudey, Steve E. 2009. “Chlorination Disinfection By-Products, Public Health Risk Tradeoffs and Me.” *Water Research* 43 (8): 2057–92. https://doi.org/10.1016/j.watres.2009.02.011.

Huong, Srei Pisei, Emmanuelle Rocher, J. D. Fourneron, Laurence Charles, Valérie Monnier, Hot Bun, and Véronique Andrieu. 2008. “Photoreactivity of the Sunscreen Butylmethoxydibenzoylmethane (DBM) under Various Experimental Conditions.” *Journal of Photochemistry and Photobiology A-Chemistry* 196: 106–12. https://doi.org/10.1016/j.jphotochem.2007.11.023.

Jacoby, Mitch. 2015. “Fish Make Their Own Sunscreens.” *Chemical & Engineering News* 93 (22). https://cen.acs.org/articles/93/i22/Fish-Make-Own-Sunscreens.html%7CFish.

Jing, H, S Jing, K Reinhard, and P Ralf. 2018. “Process for the Manufacture of Substituted 2-Cyano Cinnamic Esters. Patent WO2008089920A1.” https://worldwide.espacenet.com/publicationDetails/biblio?CC=WO&NR=2008089920&KC=&FT=E&locale=en_EP.

Jordan, Rob. 2022. “Understanding How Sunscreens Damage Coral.” *Stanford University News* (blog). May 5, 2022. https://news.stanford.edu/2022/05/05/coral-killing-sunscreens/.

Joshi, Hem, Cees Gooijer, and Gert Zwan. 2002. “Water-Induced Quenching of Salicylic Anion Fluorescence.” *Journal of Physical Chemistry A - J PHYS CHEM A* 106 (November). https://doi.org/10.1021/jp020442s.

Kariagina, Anastasia, Elena Morozova, Reyhane Hoshyar, Mark D. Aupperlee, Mitchell A. Borin, Sandra Z. Haslam, and Richard C. Schwartz. 2020. “Benzophenone-3 Promotion of Mammary Tumorigenesis Is Diet-Dependent.” *Oncotarget* 11 (28): 4465–78. https://doi.org/10.18632/oncotarget.27831.

Karlsson, Isabella, Lisa Hillerström, Anna-Lena Stenfeldt, Jerker Mårtensson, and Anna Börje. 2009. “Photodegradation of Dibenzoylmethanes: Potential Cause of Photocontact Allergy to Sunscreens.” *Chemical Research in Toxicology* 22 (11): 1881–92. https://doi.org/10.1021/tx900284e.

Kart, Jeff. 2022. “Arcaea Acquires Gadusol To Speed Development Of Natural Sunscreen That Protects You Like A Fish.” *Forbes*, August 2, 2022. https://www.forbes.com/sites/jeffkart/2022/08/02/arcaea-acquires-gadusol-to-speed-development-of-natural-sunscreen-that-protects-you-like-a-fish/?sh=564ac2a05951.

Kaupp, Gerd, M Reza Naimi-Jamal, and Jens Schmeyers. 2003. “Solvent-Free Knoevenagel Condensations and Michael Additions in the Solid State and in the Melt with Quantitative Yield.” *Tetrahedron* 59 (21): 3753–60. https://doi.org/10.1016/S0040-4020(03)00554-4.

Kimbrough, Doris R. 1997. “The Photochemistry of Sunscreens.” *Journal of Chemical Education* 74 (1): 51. https://doi.org/10.1021/ed074p51.

Krause, M., A. Klit, M. Blomberg Jensen, T. Søeborg, H. Frederiksen, M. Schlumpf, W. Lichtensteiger, N. E. Skakkebaek, and K. T. Drzewiecki. 2012. “Sunscreens: Are They Beneficial for Health? An Overview of Endocrine Disrupting Properties of UV-Filters.” *International Journal of Andrology*, May. https://doi.org/10.1111/j.1365-2605.2012.01280.x.

Kunisue, Tatsuya, Zhen Chen, Germaine M. Buck Louis, Rajeshwari Sundaram, Mary L. Hediger, Liping Sun, and Kurunthachalam Kannan. 2012. "Urinary Concentrations of Benzophenone-Type UV Filters in U.S. Women and Their Association with Endometriosis." *Environmental Science & Technology* 46 (8): 4624–32. https://doi.org/10.1021/es204415a.

La Farré, Marinel, Sandra Pérez, Lina Kantiani, and Damià Barceló. 2008. "Fate and Toxicity of Emerging Pollutants, Their Metabolites and Transformation Products in the Aquatic Environment." *Advanced MS Analysis of Metabolites and Degradation Products - II* 27 (11): 991–1007. https://doi.org/10.1016/j.trac.2008.09.010.

Lhiaubet-Vallet, Virginie, Mireia Marin, Oscar Jimenez, Olga Gorchs, Carles Trullas, and Miguel Angel Miranda. 2010. "Filter-Filter Interactions. Photostabilzation, Triplet Quenching and Reactivity with Singlet Oxygen." *Photochemical and Photobiological Sciences* 9 (4): 552–58. https://doi.org/10.1039/b9pp00158a.

Matta, Murali K, Robbert Zusterzeel, Nageswara R Pilli, Vikram Patel, Donna A Volpe, Jeffry Florian, Luke Oh, et al. 2019. "Effect of Sunscreen Application Under Maximal Use Conditions on Plasma Concentration of Sunscreen Active Ingredients: A Randomized Clinical Trial." *JAMA* 321 (21): 2082–91. https://doi.org/10.1001/jama.2019.5586.

———. 2020. "Effect of Sunscreen Application on Plasma Concentration of Sunscreen Active Ingredients: A Randomized Clinical Trial." *JAMA* 323 (3): 256–67. https://doi.org/10.1001/jama.2019.20747.

McMahon, Shannon. 2021. "These 7 Destinations Are Banning Certain Sunscreens." *Condé Nast Traveler: News & Advice*, May 11, 2021. https://www.cntraveler.com/story/these-destinations-are-banning-certain-sunscreens.

Michele, Theresa. 2021. "An Update on Sunscreen Requirements: The Deemed Final Order and the Proposed Order." FDA. https://www.fda.gov/drugs/news-events-human-drugs/update-sunscreen-requirements-deemed-final-order-and-proposed-order.

Mitchelmore, Carys L., Ke He, Michael Gonsoir, Ethan Hain, Andrew Heyes, Cheryl Clark, Rick Younger, et al. 2019. "Occurrence and Distribution of UV-Filters and Other Anthropogenic Contaminants in Coastal Surface Water, Sediment, and Coral Tissue from Hawaii." *Science of The Total Environment* 670 (June): 398–410. https://doi.org/10.1016/j.scitotenv.2019.03.034.

Mori, Shoko, and Steven Q. Wang. 2021. "50 - Sunscreens." *Comprehensive Dermatologic Drug Therapy (Fourth Edition)*, 565-575.e2. https://doi.org/10.1016/B978-0-323-61211-1.00050-4.

NCBI. 2022a. "PubChem Compound Summary for CID 4632, Oxybenzone." https://pubchem.ncbi.nlm.nih.gov/compound/Oxybenzone.

———. 2022b. "PubChem Compound Summary for CID 8362, Homosalate." https://pubchem.ncbi.nlm.nih.gov/compound/Homosalate#section=Chemical-and-Physical-Properties.

———. 2022c. "PubChem Compound Summary for CID 8364, 2-Ethylhexyl Salicylate." https://pubchem.ncbi.nlm.nih.gov/compound/2-Ethylhexyl-salicylate#section=Chemical-and-Physical-Properties.

———. 2022d. “PubChem Compound Summary for CID 22571, Octocrylene.” https://pubchem.ncbi.nlm.nih.gov/compound/Octocrylene#section=Chemical-and-Physical-Properties.
———. 2022e. “PubChem Compound Summary for CID 51040, Avobenzone.” https://pubchem.ncbi.nlm.nih.gov/compound/Avobenzone.
———. 2022f. “PubChem Compound Summary for CID 195955, Gadusol.” https://pubchem.ncbi.nlm.nih.gov/compound/Gadusol.
———. 2022g. “PubChem Compound Summary for CID 5280343, Quercetin.” https://pubchem.ncbi.nlm.nih.gov/compound/Quercetin.
———. 2022h. “PubChem Compound Summary for CID 5354031, Ubiquinone.” https://pubchem.ncbi.nlm.nih.gov/compound/Ubiquinone#section=Chemical-and-Physical-Properties.
———. 2022i. “PubChem Compound Summary for CID 5355130, Octinoxate.” https://pubchem.ncbi.nlm.nih.gov/compound/Octinoxate#section=Computed-Properties.
———. 2022j. “PubChem Compound Summary for CID 6439771.” https://pubchem.ncbi.nlm.nih.gov/compound/6439771.
Nguyen, V. P. Thinh, Jon D. Stewart, Irina Ioannou, and Florent Allais. 2021. “Sinapic Acid and Sinapate Esters in Brassica: Innate Accumulation, Biosynthesis, Accessibility via Chemical Synthesis or Recovery From Biomass, and Biological Activities.” *Frontiers in Chemistry*, May. https://doi.org/10.3389/fchem.2021.664602.
NOAA. 2021a. “How Does Climate Change Affect Coral Reefs?” https://oceanservice.noaa.gov/facts/coralreef-climate.html.
———. 2021b. “What Is Coral Bleaching?” https://oceanservice.noaa.gov/facts/coral_bleach.html.
NPS. n.d. “Protect Yourself, Protect The Reef!” https://cdhc.noaa.gov/_docs/Site%20Bulletin_Sunscreen_final.pdf.
Osterloff, Emily. n.d. “What Is Ocean Acidification?” National History Museum. https://www.nhm.ac.uk/discover/what-is-ocean-acidification.html.
Peinado, F.M., O. Ocón-Hernández, L.M. Iribarne-Durán, F. Vela-Soria, A. Ubiña, C. Padilla, J.C. Mora, et al. 2021. “Cosmetic and Personal Care Product Use, Urinary Levels of Parabens and Benzophenones, and Risk of Endometriosis: Results from the EndEA Study.” *Environmental Research* 196 (May): 110342. https://doi.org/10.1016/j.envres.2020.110342.
Peyrot, Cédric, Matthieu Mention, Fanny Brunissen, and Florent Allais. 2020. “Sinapic Acid Esters: Octinoxate Substitutes Combining Suitable UV Protection and Antioxidant Activity.” *Antioxidants* 9 (9). https://doi.org/10.3390/antiox9090782.
Richardson Susan D., DeMarini David M., Kogevinas Manolis, Fernandez Pilar, Marco Esther, Lourencetti Carolina, Ballesté Clara, et al. 2010. “What’s in the Pool? A Comprehensive Identification of Disinfection By-Products and Assessment of Mutagenicity of Chlorinated and Brominated Swimming Pool Water.” *Environmental Health Perspectives* 118 (11): 1523–30. https://doi.org/10.1289/ehp.1001965.

Richardson, Susan D., Michael J. Plewa, Elizabeth D. Wagner, Rita Schoeny, and David M. DeMarini. 2007. "Occurrence, Genotoxicity, and Carcinogenicity of Regulated and Emerging Disinfection by-Products in Drinking Water: A Review and Roadmap for Research." *The Sources and Potential Hazards of Mutagens in Complex Environmental Matrices - Part II* 636 (1): 178–242. https://doi.org/10.1016/j.mrrev.2007.09.001.

Rioux, Benjamin, Matthieu Mention, Jimmy Alacran, Temitrope T. Abiola, Cédric Peyrot, and Fanny Brunissen. 2022. "Sustainable Synthesis, in Silico Evaluation of Potential Toxicity and Environmental Fate, Antioxidant and UV-Filtering/Photostability Activity of Phenolic-Based Thiobarbituric Derivatives." *Green Chemistry Letters and Reviews* 15 (1): 116–27. https://doi.org/10.1080/17518253.2021.2022219.

Rooney, John, Natalia Ryan, Jie Liu, René Houtman, Rinie van Beuningen, Jui-Hua Hsieh, Gregory Chang, Shiuan Chen, and J. Christopher Corton. 2021. "A Gene Expression Biomarker Identifies Chemical Modulators of Estrogen Receptor α in an MCF-7 Microarray Compendium." *Chemical Research in Toxicology* 34 (2): 313–29. https://doi.org/10.1021/acs.chemrestox.0c00243.

Russo, J.P., A. Ipiña, J.F. Palazzolo, A.B. Cannavó, R.D. Piacentini, and B. Niklasson. 2018. "Dermatitis Por Contacto Fotoalérgica a Protectores Solares Con Oxibenzona En La Plata, Argentina." *Actas Dermo-Sifiliográficas* 109 (6): 521–28. https://doi.org/10.1016/j.ad.2018.02.011.

Salvador, Amparo, Alberto Chisvert, A. Camarasa, M. C. Pascual-MartÍ, and J. G. March. 2001. "Sequential Injection Spectrophotometric Determination of Oxybenzone in Lipsticks." *Analyst* 126 (8): 1462–65. https://doi.org/10.1039/b103497a.

Sánchez-Quiles, David, and Antonio Tovar-Sánchez. 2014. "Sunscreens as a Source of Hydrogen Peroxide Production in Coastal Waters." *Environmental Science & Technology* 48 (16): 9037–42. https://doi.org/10.1021/es5020696.

Santos, A Joel M, Margarida S Miranda, and Joaquim C. G. Esteves Da Silva. 2012. "The Degradation Products of UV Filters in Aqueous and Chlorinated Aqueous Solutions." *Water Research* 46 (10): 3167–76. https://doi.org/10.1016/j.watres.2012.03.057.

Scalia, Santo, and Mateo Mezzena. 2010. "Photostabilization Effect of Quercetin on the UV Filter Combination, Butyl Methoxydibenzoylmethane–Octyl Methoxycinnamate." *Photochemical and Photobiological Sciences* 86 (2): 273–78. https://doi.org/10.1111/j.1751-1097.2009.00655.x.

Schijndel, Jack van, Luiz Alberto Canalle, Dennis Molendijk, and Jan Meuldijk. 2017. "The Green Knoevenagel Condensation: Solvent-Free Condensation of Benzaldehydes." *Green Chemistry Letters and Reviews* 10 (4): 404–11. https://doi.org/10.1080/17518253.2017.1391881.

Schneider, Samantha L., and Henry W Lim. 2019. "Review of Environmental Effects of Oxybenzone and Other Sunscreen Active Ingredients." *Journal of the American Academy of Dermatology* 80 (1): 266–71. https://doi.org/10.1016/j.jaad.2018.06.033.

Scinicariello Franco and Buser Melanie C. 2016. “Serum Testosterone Concentrations and Urinary Bisphenol A, Benzophenone-3, Triclosan, and Paraben Levels in Male and Female Children and Adolescents: NHANES 2011–2012.” *Environmental Health Perspectives* 124 (12): 1898–1904. https://doi.org/10.1289/EHP150.

Shaath, Nadim A. 2010. “Ultraviolet Filters.” *Photochemical and Photobiological Sciences* 9 (4): 464–69. https://doi.org/10.1039/B9PP00174C.

Smith, Anna. 2018. “Is Your Sunscreen Killing the Coral Reef?” *Ocean Conservancy Blog: Ocean Currents* (blog). May 24, 2018. https://oceanconservancy.org/blog/2018/05/24/sunscreen-killing-coral-reef/.

Spitz, Guilherme A., Cristiane M. Furtado, Mauro Sola-Penna, and Patricia Zancan. 2009. “Acetylsalicylic Acid and Salicylic Acid Decrease Tumor Cell Viability and Glucose Metabolism Modulating 6-Phosphofructo-1-Kinase Structure and Activity.” *Biochemical Pharmacology* 77 (1): 46–53. https://doi.org/10.1016/j.bcp.2008.09.020.

State of Hawaii. 2018. *Relating to Water Pollution*. https://www.capitol.hawaii.gov/session2018/bills/SB2571_.HTM.

Stien, Didier, Fanny Clergeaud, Alice M. S. Rodrigues, Karine Lebaron, Rémi Pillot, Pascal Romans, Sonja Fagervold, and Philippe Lebaron. 2018. “Metabolomics Reveal That Octocrylene Accumulates in Pocillopora Damicornis Tissues as Fatty Acid Conjugates and Triggers Coral Cell Mitochondrial Dysfunction.” *Analytical Chemistry* 91 (1): 990–95. https://doi.org/10.1021/acs.analchem.8b04187.

Todorov, G. n.d. “Chemical UVB+UVA Sunscreen/Sunblock: Octocrylene.” SmartSkinCare.Com. http://www.smartskincare.com/skinprotection/sunblocks/sunblock_octocylene.html.

Tournas, Joshua A., Fu-Hsiung Lin, James A. Burch, M Angelia Selim, Nancy A Monteiro-Riviere, Jan E. Zielinski, and Sheldon R. Pinnell. 2006. “Ubiquinone, Idebenone, and Kinetin Provide Ineffective Photoprotection to Skin When Compared to a Topical Antioxidant Combination of Vitamins C and E with Ferulic Acid.” *Journal of Investigative Dermatology* 126 (5): 1185–87. https://doi.org/10.1038/sj.jid.5700232.

Vaghari, Hamideh, Roholah Vaghari, Hoda Jafarizadeh-Malmiri, and Aydin Berenjian. 2016. “Coenzyme Q10 and Its Effective Sources.” *American Journal of Biochemistry and Biotechnology* 12 (4): 214–19. https://doi.org/10.3844/ajbbsp.2016.214.219.

Vuckovic, Djordje, Amanda L. Tinoco, Lorraine Ling, Christian Renicke, John R. Pringle, and William A. Mitch. 2022. “Conversion of Oxybenzone Sunscreen to Phototoxic Glucoside Conjugates by Sea Anemones and Corals.” *Science* 376 (6593): 644–48. https://doi.org/10.1126/science.abn2600.

Wang, Steven Q., Mark E. Burnett, and Henry W. Lim. 2011. “Safety of Oxybenzone: Putting Numbers Into Perspective.” *Archives of Dermatology* 147 (7): 865–66. https://doi.org/10.1001/archdermatol.2011.173.

Wang, Ying, and Siegfried Hekimi. 2020. "Micellization of Coenzyme Q by the Fungicide Caspofungin Allows for Safe Intravenous Administration to Reach Extreme Supraphysiological Concentrations." *Redox Biology* 36 (September): 101680. https://doi.org/10.1016/j.redox.2020.101680.

WHO. 2016. "Radiation: Ultraviolet (UV) Radiation," March.

Wu, Haiyou, Zhangfeng Zhong, Sien Lin, Chuqun Qiu, Peitao Xie, Simin Lv, Liao Cui, and Tie Wu. 2020. "Coenzyme Q10 Sunscreen Prevents Progression of Ultraviolet-Induced Skin Damage in Mice." Edited by Davinder Parsad. *BioMed Research International* 2020 (August): 9039843. https://doi.org/10.1155/2020/9039843.

Yan, Saihong, Mengmeng Liang, Rui Chen, Xiangsheng Hong, and Jinmiao Zha. 2020. "Reproductive Toxicity and Estrogen Activity in Japanese Medaka (Oryzias Latipes) Exposed to Environmentally Relevant Concentrations of Octocrylene." *Environmental Pollution* 261 (June): 114104. https://doi.org/10.1016/j.envpol.2020.114104.

Yazar, Selma, and Simge Kara Ertekin. 2020. "Assessment of the Cytotoxicity and Genotoxicity of Homosalate in MCF-7." *Journal of Cosmetic Dermatology* 19 (1): 246–52. https://doi.org/10.1111/jocd.12973.

Zdravković, Tanja Prunk, Bogdan Zdravković, Marko Zdravković, Barbara Dariš, Mojca Lunder, and Polonca Ferk. 2019. "In Vitro Study of the Influence of Octocrylene on a Selected Metastatic Melanoma Cell Line." *Italian Journal of Dermatology and Venereology* 154 (2): 197–204. https://doi.org/10.23736/S0392-0488.17.05616-4.

Zhang, Qiuya Y., Xiaoyan Y. Ma, Xiaochang C. Wang, and Huu Hao Ngo. 2016. "Assessment of Multiple Hormone Activities of a UV-Filter (Octocrylene) in Zebrafish (Danio Rerio)." *Chemosphere* 159 (September): 433–41. https://doi.org/10.1016/j.chemosphere.2016.06.037.

Supplementary Materials

Table 1. *Summary of Current UV Filters*

Name	Max Absorbance	LogP	Summary
Oxybenzone	286 nm (UVB)	3.6	Potential endocrine disruption; highest skin absorption; toxic metabolites formed in coral; triggering coral viruses and bleaching; genotoxin
Octinoxate	311 nm (UVB)	5.3	Coral bleaching; endocrine disruption; fossil-fuel intensive production
Avobenzone	357 nm (UVA)	4.8	Most used UVA filter; harmful reaction pathways of photodegradation products and keto form; photoallergen; cytotoxin
Octocrylene	303 (weak UVB)	7.1	Photostabilizer for avobenzone; photoallergen; generation of free radicals; contaminated with carcinogenic benzophenone

Name	Max Absorbance	LogP	Summary
Homosalate	306 (UVB)	5	Pass skin barriers at levels exceeding FDA cutoffs; endocrine disruption; impact MCF-7 (breast cancer) cells; bioaccumulation
Octisalate	305 (UVB)	5.7	Pass skin barriers at levels exceeding FDA cutoffs; endocrine disruption; bioaccumulation

Table 2. *Summary of Proposed Compounds*

Name	Max Absorbance	LogP	Summary
Ubiquinone	275 nm (UVB)	19.4	Antioxidant; stabilize avobenzone; quench radicals formed during UV-radiation; yellow pigmentation; very low likelihood of toxicity
Quercetin	240-280 nm, 340-440 nm	1.5	Antioxidant; stabilization of avobenzone and octinxate; orange/yellow pigmentation, potential for tinted cosmetics; very low likelihood of toxicity
Sinapate and Sinapic Acid Esters	UVA/UVB	Varies	Plant sunscreens; antioxidant properties; potential photostabilization of other compounds; easily derived or synthesized; very low likelihood of toxicity
DHDES	~335 nm	--	Ethyl sinapate dimer with greater absorbance properties; antioxidant; potential photostabilization of other compounds; octinoxate substitute; unknown photodegradation products
Thiobarbituric Acid Derivatives	399-445 nm (UVA)	Varies	Low loss of absorbance abilities after extended UV exposure; antioxidant; easily synthesized; very low likelihood of toxicity

A Computational Study on the Mechanism and the Regio- and Stereoselectivity of Metal-mediated Nucleophilic Addition to *Gem*-difluoroallene

Michael Li

Author Background: *Michael Li grew up in Canada and currently attends St. George's School in Vancouver, British Columbia in Canada. His Pioneer research concentration was in the field of chemistry and titled "Computational Organic Chemistry."*

Abstract

Allenes that contain two geminal fluorine, known as *gem*-difluoroallenes, gained significant research attention in recent years due to their unique chemical reactivity. One type of reaction that *gem*-difluoroallenes can undergo is nucleophilic addition mediated by select metals. However, prior studies have focused on the experimental development for nucleophilic addition and have provided limited computational insight into the reaction mechanism and chemical selectivity. In this study, computational data were gathered from calculations done on Gaussian09. These calculations were done at ωB97X-D level with a mixed basis set of 6-311G**(C,H), 6-311+G*(F), and LANL2TZ(f) for the metals copper, silver, and gold. Specifically, the preference for the E isomer product and α-addition, as well as the mechanism behind the nucleophilic addition of AgF to *gem*-difluoroallene, were analyzed. One significant finding of this study is that Ag^+ coordinates with one of the double bonds in the allene structure, specifically on the non-fluorinated side. Subsequent nucleophilic attack of F^- on the α carbon produces the lowest energy pathway to forming the addition product with a trifluoromethylated molecule. It is this step that leads to the high stereoselectivity of the reaction as there can be unfavorable interactions between the nucleophile and substituted groups on *gem*-difluoroallene. While the specific, examined reaction is a simplified system, the computational data in this study serve as a more quantitative explanation for experimental findings concerning nucleophilic addition to *gem*-difluoroallenes and provide insight into the reaction mechanism.

1. Introduction

Fluorine is known for its high electronegativity and has useful applications in the pharmaceutical and agrochemical industries.[1] CF_2-containing molecules have been of particular interest, most notably *gem*-difluoroalkenes where two geminal fluorine atoms are attached to one end of an alkene.[2] Molecules with *gem*-difluoroalkene

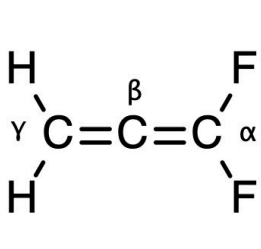
moieties are potentially useful intermediates in organic syntheses, with many new reactions being developed that involve the cleavage and functionalization of the carbon-fluorine bonds.[2] They have been shown to participate in reactions including S_NV reactions,[3-5] nucleophilic intramolecular cyclization reactions,[6-8] and metal-mediated cross-coupling and cyclization reactions.[9-17]

While much research has gone into the synthesis and utilization of *gem*-difluoroalkenes, another similar CF_2-containing structural moiety called *gem*-difluoroallenes, or 1,1-difluoroallene, has shown unique reactivity and selectivity. *Gem*-difluoroallenes contain the same geminal fluorine atoms at one end of an allene which involves two adjacent, or cumulative, double bonds with orthogonal π-systems. Prior research has exhibited the possibility for cycloaddition and hydrometalation with *gem*-difluoroallene,[18-22] but this study will primarily investigate nucleophilic addition through the use of metal reagents. This type of reaction with *gem*-difluoroallene has been shown to produce high yields under particular conditions. However, there lacks a clear understanding of the various detailed mechanisms and the cause for regioselectivity and stereoselectivity when utilizing different reagents. Therefore, a computation study on nucleophilic addition to *gem*-difluoroallene can help explain previous experimental results and potentially provide a framework for developing new reaction pathways for this unique fluorine-containing group.

2. Overview

2.1. *Gem*-difluoallene Properties

The introduction of two geminal fluorine atoms on the α carbon of an allene allows the molecule of *gem*-difluoroallene to have unique properties. For one, the charges on each carbon differ significantly (see Figure 1) due to the addition of highly electronegative atoms on one end.



	C(α)	C(β)	C(γ)
Electrostatic:	+0.276	−0.085	−0.333
Mulliken:	+0.349	+0.221	−0.490

Figure 1. *Electrostatic and Mulliken charges on each atom of gem-difluoroallene calculated using B3LYP/6-31G*.*

Although the electropositive α carbon may suggest a likely position for nucleophilic addition, it is important to note that the concentrated negative charges on the geminal fluorine atoms may lead to undesirable attack from certain anionic nucleophiles. Furthermore, organic reactions are also governed by orbital

interactions. Figure 2 shows the frontier molecular orbitals for *gem*-difluoroallene.

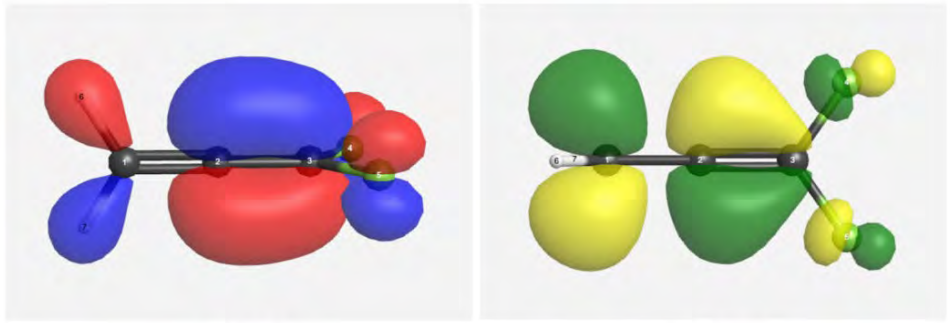


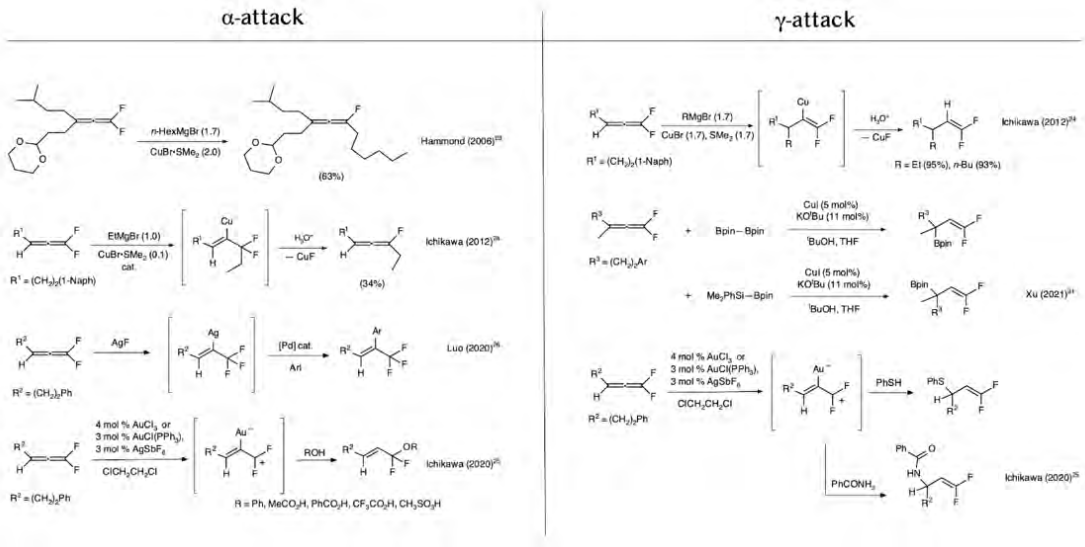
Figure 2. *The calculated frontier molecular orbitals (the HOMO and LUMO) for gem-difluoroallene.*

This molecule features unique dynamic orbital and electrostatic interactions, and they are what allow for regioselective nucleophilic addition to *gem*-difluoroallene.

2.2. Previous Experimental Work

Hammond (2006) reported the possibility of SNV reaction on *gem*-difluoroallene through nucleophilic attack on the α carbon.[23] This is depicted in Table 1. The group utilized copper as the metal to mediate the reaction and had a yield of 63%.

Table 1. Previous experimental reports of α or γ nucleophilic attack on gem-difluoroallene. The percentages inside parentheses indicate yield.



Then in 2012, Ichikawa et al. performed a similar experiment with copper but utilizing a *gem*-difluroallene molecule with a more accessible γ carbon along with a variety of different reactants and concentrations.[24] Under a catalytic amount of the Cu(I) species, the reaction seemed to favor α attack, though it produced low yield. However, with a larger amount of copper, the group reported yields of more than 90% favoring γ-attack, which resulted in the formation of *gem*-difluoroalkenes. No reaction was observed with MeLi or EtMgBr, while $ZnEt_2$ produced a low 12% yield of the α addition-elimination product.

The same authors in 2020 provided an experimental report of a catalyzed nucleophilic addition to *gem*-difluoroallene through the use of gold.[25] Interestingly, there was a high degree of regioselectivity between different nucleophiles, with *O*-nucleophiles undergoing α-addition and *N*- and *S*-nucleophiles undergoing γ-addition. Ichikawa's team suggested an intermediate where the gold catalyst attaches to the β carbon and forms a positive charge on the α carbon. They utilize this intermediate structure in their computational calculations to suggest why there was high E-stereoisomerism observed in the products.

In the same year, Luo (2020) showed that the addition of AgF and subsequently a palladium catalyst with an aryl halide can produce a trifluoromethylated product with the aryl group attached to the labelled β carbon where the silver ion would have been attached, according to their proposed intermediate step.[26] Other studies have incorporated AgF in forming a trifluoromethylated product, but the focus was on *gem*-difluroalkenes.[27-29] In particular, one study suggested a two-step mechanism where the fluorine anion nucleophile attacks *gem*-difluoroalkene, forming a carbanion intermediate, followed by electrophilic halogenation.[28] This pathway is different compared to the proposed mechanism by Ichikawa's group in 2020, suggesting instead a cationic intermediate, although they investigated a different molecule and metal catalyst. Ultimately, a computationally centered analysis would provide a clearer picture of the mechanism of nucleophilic addition to *gem*-difluoroallene and the cause of the regioselectivity and stereoselectivity.

3. Methodology

For this computational study, the reaction between *gem*-difluroallene with AgF, which forms a trifluoromethylated product, will be primarily investigated, and the findings can be further applied to more complicated systems. The reason for this choice is that fluorine is one of the simplest nucleophiles, and previous experiments have shown that AgF almost solely undergoes α-addition. Furthermore, unless specified, calculations are done on an unsubstituted gem-difluoroallene molecule.

These calculations were primarily done on Gaussian09 at ωB97X-D level of theory. Considering the accuracy required and also the computational cost, a mixed basis set approach was chosen to be 6-311G** (C, H), 6-311+G* (F), and LANL2TZ(f) (Ag). In terms of the solvent model, acetonitrile was selected. Finally, for transition states or intermediates, added components during steps were taken into account when calculating the relative energies in energy profiles.

4. Results and Discussion

4.1. Calculated Mechanisms of α-addition

Luo (2020) utilized AgF dissolved in an acetonitrile solution to react with gem-difluoroallene through α-addition of a fluorine anion nucleophile.[26] However, the mechanism concerning the regiospecific and stereospecific coordination of F– and Ag+ to the molecule was not described.

4.1.1 Plausible Reaction Pathways

A possible pathway is first having the F^- nucleophile attack *gem*-difluoroallene, forming an anionic intermediate, and then subsequently having an Ag^+ ion coordinate (Pathway 1). Alternatively, free Ag^+ ions can coordinate to *gem*-difluoroallene first, which can then foster nucleophilic attack from F^- (Pathway 2).

4.1.2 Coordination of the Silver Ion in Pathway 2

Before analyzing the two pathways together, there warrants an investigation into the intermediate bonding structure in Pathway 2. Specifically, the coordination of the Ag+ ion to gem-difluoroallene lowers the energy of the molecule. Interestingly, this coordination seems to involve two carbons and the double bond between them (see Figure 3). Further, Ag+ can coordinate to either of the two double bonds in the allene structure.

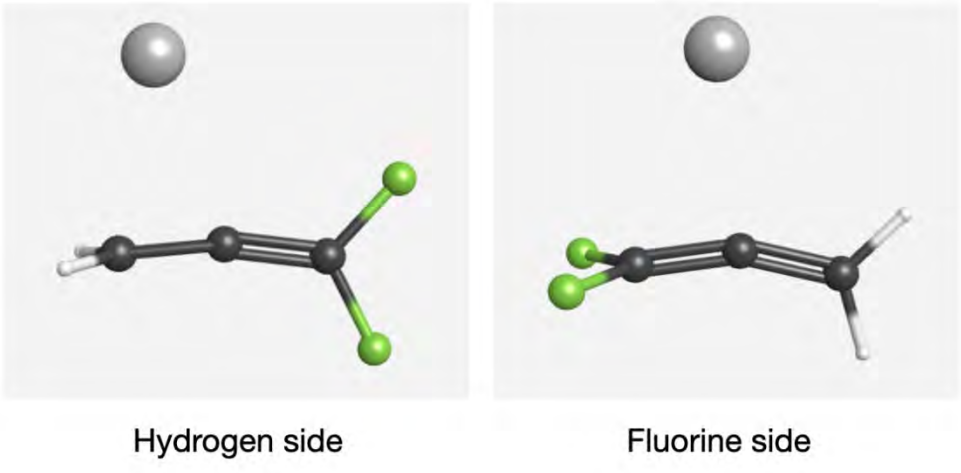


Figure 4. *Optimized geometries for the coordination of Ag^+ to gem-difluoroallene.*

For the coordination to the hydrogen side, the Ag–C bond lengths are both 2.495Å to the β and γ carbons, and the bond length between these two carbons is 1.321Å, which is slightly longer than the bond length of 1.301Å without the silver ion interaction. Coordination to the fluorine side produced Ag–C bond lengths of 2.849Å and 2.512Å to the α and β, respectively, while the bond length between these two carbons is 1.315Å compared to 1.292Å without the coordination. The larger bond

lengths for coordination to the fluorine side are likely due to the repulsive interaction between the electropositive α carbon and the positively charged silver ion, leading to a less tightly bound structure.

This interaction has been seen with Ag^+ ions and olefins.[30] The coordination of the silver ion takes place primarily due to the orbital interaction of C=C π bond and the empty s-orbital of Ag^+, but also, to an extent, back-bonding electron donation from the Ag^+ 4d orbital to the C=C π* orbital. Because of the observed position of the silver ion bonded to each double bond on *gem*-difluoroallene, a similar interaction seems to take place as well, which will be investigated further in section 4.4.

This finding is one potential explanation for unsuccessful attempts to find a carbocation intermediate using ^{19}F NMR in Ichicawa's 2020 study on gold catalysts. This proposed coordination maintains most of the positive charge on the metal ion. For instance, in the two structures with Ag^+ coordinating on the hydrogen or fluorine side shown in Figure 3, the Mulliken charges on the α carbon are +0.36 and +0.46, respectively, while the charges on the silver atoms are both +0.86.

In solution, the coordination to the hydrogen side is about 2.4 kcal/mol lower in energy compared to the fluorine side, according to molecular energy calculations of the complexes. However, both lower the energy of *gem*-difluoroallene by 5.1 kcal/mol and 2.7 kcal/mol, respectively. Due to these two possibilities, Pathway 2 will be separated into Pathway 2A and Pathway 2B, corresponding to Ag^+ first coordinating to the hydrogen side or the fluorine side, respectively.

4.1.3 Analysis of the Plausible Pathways

Figure 4 below depicts the energy profile of each of the three identified pathways from DFT calculations. Pathway 1 had an energy demand of 11.1 kcal/mol for the F^- nucleophilic attack, which achieved an anionic intermediate **INT1P1**. The attraction between the electropositive silver ions and **INT1P1** afforded product **P1** at significantly lower relative energy. The coordination of Ag^+ to the fluorine side, forming the intermediate **INT1P2B** in Pathway 2B was shown to lower the energy of the molecule by 2.7 kcal/mol. Subsequent nucleophilic addition of the fluorine anion produced the product, notably as a seemingly barrierless reaction. On the other hand, the coordination of Ag^+ to form **INT1P2A** lowered the energy of *gem*-difluoroallene by a larger amount of 5.3 kcal/mol. The nucleophilic addition of the fluorine anion then produced a transition state, **TS1P2A**, with an energy barrier of 1.6 kcal/mol before generating the product.

Of the three pathways, Pathway 2A and Pathway 2B are the more favorable, as the energy demand for Pathway 1 is relatively high. Comparing the two, it is evident that **INT1P2A** and **TS1P2A** remain lower in energy than **INT1P2B**. This suggests that Pathway 2A is the likely mechanism behind the addition of AgF to *gem*-difluoroallene, considering the energy profile of the possible pathways.

Figure 5 shows the calculated geometries for **TS1P2A** and **P1**. An important difference is the position of the silver ion. In **P1**, it is on the same plane as the movement of the added fluorine, while in **TS1P2A**, it appears to be orthogonal to that plane. Evidently, there is a rotation of the silver atom after the transition state to reach its final, stable position in **P1**.

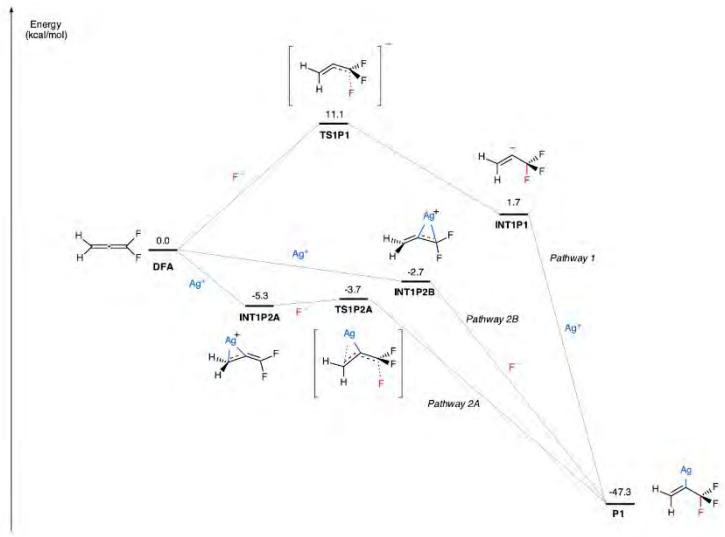


Figure 4. *Energy profiles for the possible pathways 1, 2A, and 2B.*

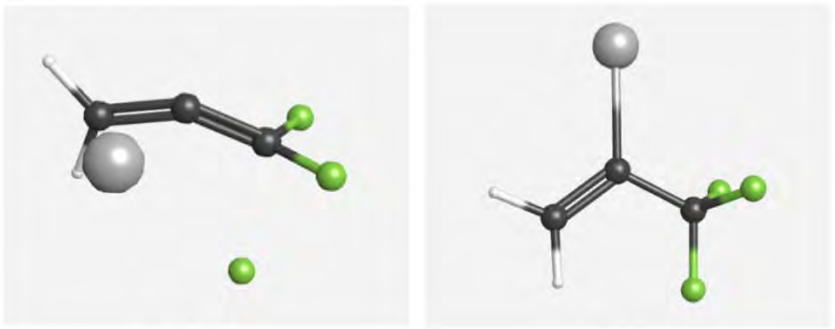


Figure 5. *Optimzied geometries of TS1P2A and P1.*

4.2. γ-addition

After the addition of the silver ion to form **INT1P2A**, γ-addition is another potential pathway for the fluorine nucleophilic attack, which will be referred to as Pathway 3. Figure 6 showcases the calculated energy profile for Pathway 3 in comparison to pathway 2A.

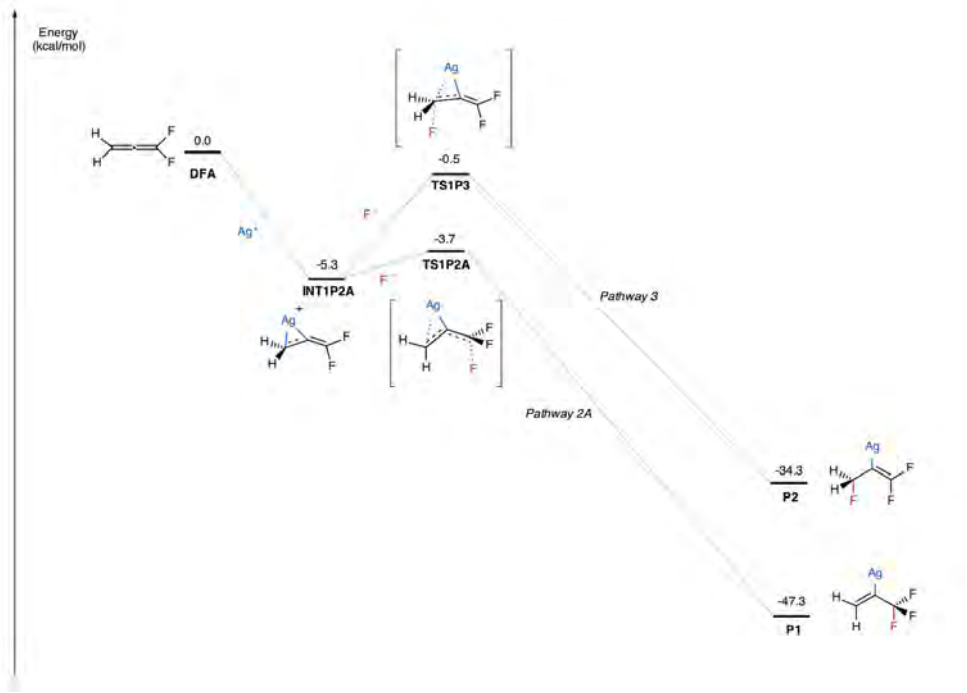


Figure 6. *Energy Profiles for the Possible Pathways 2A and 3.*

While both pathways share the same intermediate structure **INT1P2A**, the activation energy for Pathway 3 was found to be 3.2 kcal/mol higher. Furthermore, the calculated energy of **P2** was 13.0 kcal/mol higher in energy than **P1**. This energy profile indicates that Pathway 2A is thermodynamically and kinetically more favorable compared to Pathway 3, which is quantitatively why the addition of AgF to gem-difluoroallene almost exclusively forms the trifluoromethylated product through α-addition.

4.3. Explanation for E-Stereoselectivity

Previous studies have shown that nucleophilic α-addition to *gem*-difluoroallene highly favors the E product.[25,26,31] For the addition of AgF specifically, the E/Z ratio was >30/1.[26] Ichikawa's 2020 study on gold catalysts presented calculations to explain this E-stereoselectivity. Using a methyl-substituted *gem*-difluoroallene molecule, the group found that, using their formulated intermediate structure, the intermediate carbocation precursor to the Z product was 2-3 kcal/mol higher than the E product. They also suggested that it was the unfavorable Me-CF_2 interaction that made the Z precursor less stable than the E precursor.[25]

However, as seen in Figure 3 with the different proposed intermediate structure, the C-C-C bond angle is relatively large, and the fluorine and hydrogens exist on essentially orthogonal planes. This indicates that the interaction between

any substituted group with CF_2 is likely minimal. Nevertheless, calculations were done to observe the relative energies of E and Z intermediate structures with this interaction in mind. Because selectivity would not be determined by this intermediate step if the silver ion bonded to the hydrogen side due to an element of symmetry, the energies of substituted *gem*-difluoroallene molecules with the silver ion bonded to the fluorine side were first investigated, as shown in Figure 7.

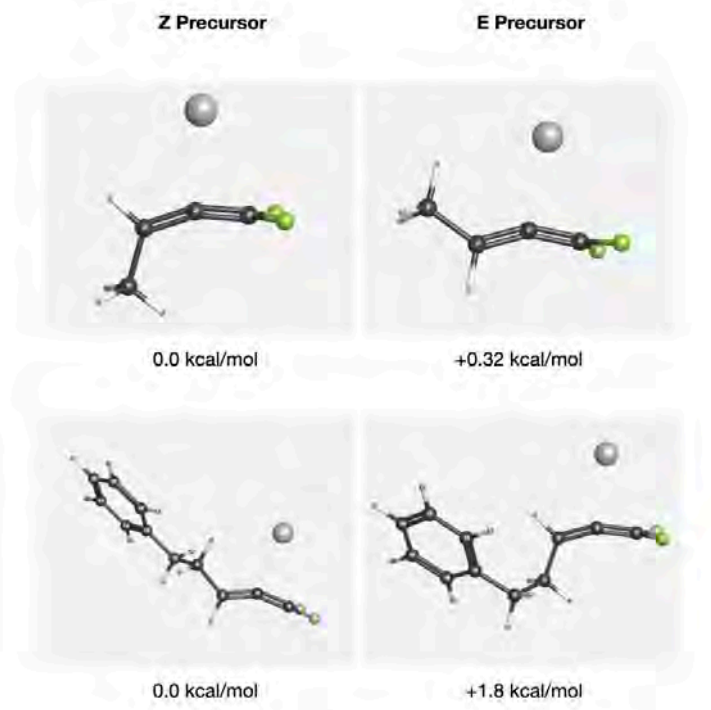


Figure 7. *Relative energies of E and Z precursors of different substituted gem-difluoroallene molecules.*

Evidently, both Z precursors are more energetically stable. Since the coordination of Ag^+ to the fluorine side would fix the direction of nucleophilic α attack of the fluorine anion, Ag^+ cannot be bonded on the fluorinated side, meaning that stereoselectivity is not determined by the metal ion coordination step. Otherwise, these energy calculations would conflict with experimental results where the E stereoisomer is preferred.

Therefore, it can be inferred that E, Z-stereoselectivity must be determined in the nucleophilic attack step with Ag^+ bonded on the hydrogen side. Since there are two directions where the fluorine can attack, the side that forms the E product will be referred to as the E face, and the side that forms the Z product will be referred to as the Z face. When the fluorine nucleophile approaches the intermediate from the Z face, there appears to be an unfavorable interaction with the substituted group. The impact of this steric hindrance is especially seen in the final products shown in Figure 8, which have a methyl group as the substituted group.

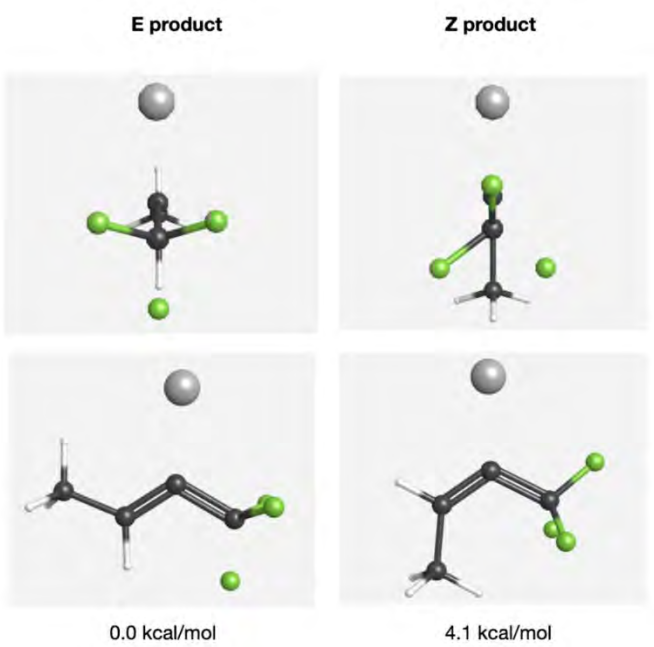


Figure 8. *Relative energies of optimized geometries of the E and Z product of methyl-substituted gem-difluoroallene reacting with AgF. The α carbons and the fluorine nucleophiles share a bond that is not visually represented.*

For the Z product, there is a rotation of the C–CF_3 bond so that the methyl group is gauche to two of the fluorine atoms. This rotation is not seen when forming the E product. It becomes apparent that the formation of the E product is more favorable than the formation of the Z product due to steric hindrance caused by the substituted group. This matches the findings of previous experimental studies.

4.4. Analyzing and comparing the cause for regioselectivity from different d^9 metals coordinating with *gem*-difluoroallene

Copper and gold have been shown to assist in nucleophilic addition to *gem*-difluoroallene.[24,25,31] Therefore, DFT calculations were conducted on copper and gold to analyze the intermediate structures formed with the double bond on the hydrogen side of *gem*-difluoroallene in acetonitrile solution. Both were shown to form a similar complex as Ag^+ (see Figure 9).

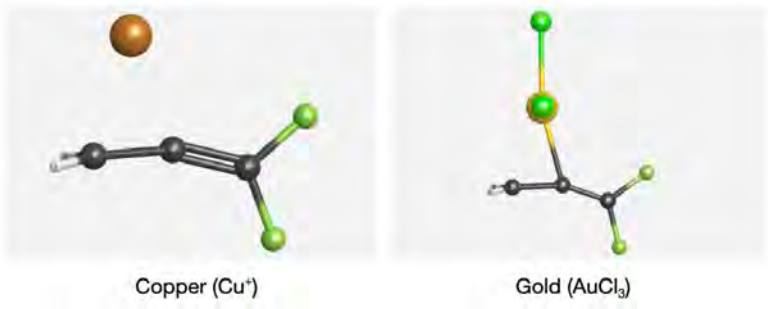


Figure 9. *Optimized geometries for Cu^+ and $AuCl_3$ complexes with gem-difluoroallene at ωB97X-D level of theory in acetonitrile solvent. Both structures were calculated using mixed basis sets: 6-311G**(C,H), 6-311+G*(F), LANL2TZ(f)(Cu) and 6-311G**(C, H), 6-311+G*(F, Cl), LANL2TZ(f)(Au) respectively.*

The interactions that form these structures were examined by looking at the bonding molecular orbitals. Considering the orientation of the molecular orbital and placement of Ag^+ in the optimized structure, it seemed most reasonable that the HOMO-1 orbital, shown in Figure 10, is primarily involved in this interaction.

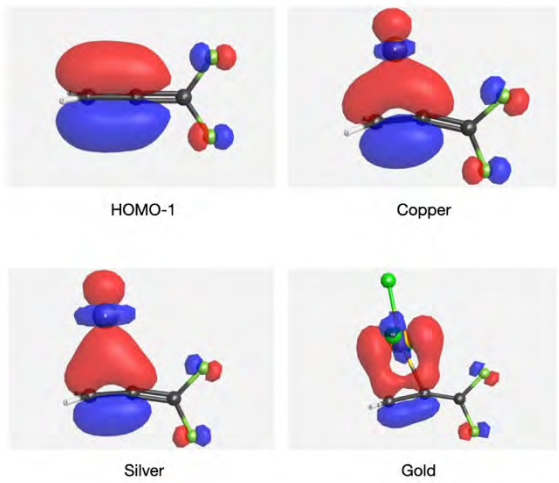


Figure 10. *Molecular orbital shapes for the HOMO-1 orbital on gem-difluoroallene and one of the bonding orbitals for each of the metal complexes.*

It appears that the HOMO-1 orbital interacts with the d_{z^2} orbital of the metals head-on for copper and silver, but for gold, the two orbitals combine in a sideways overlap manner. Further, d-orbitals with a z-component and proper energies can be involved in slight back-bonding interactions with *gem*-difluoroallene shown below.

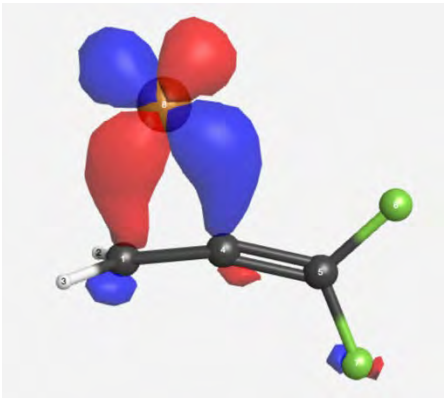


Figure 11: *Back-bonding interaction with copper and gem-difluoroallene.*

To investigate the regioselectivity in the nucleophilic attack step, the LUMO and LUMO+1 orbital shapes were analyzed and shown in Figure 12. The level of theory, solvent model, and basis sets were the same as in previous calculations for each metal complex.

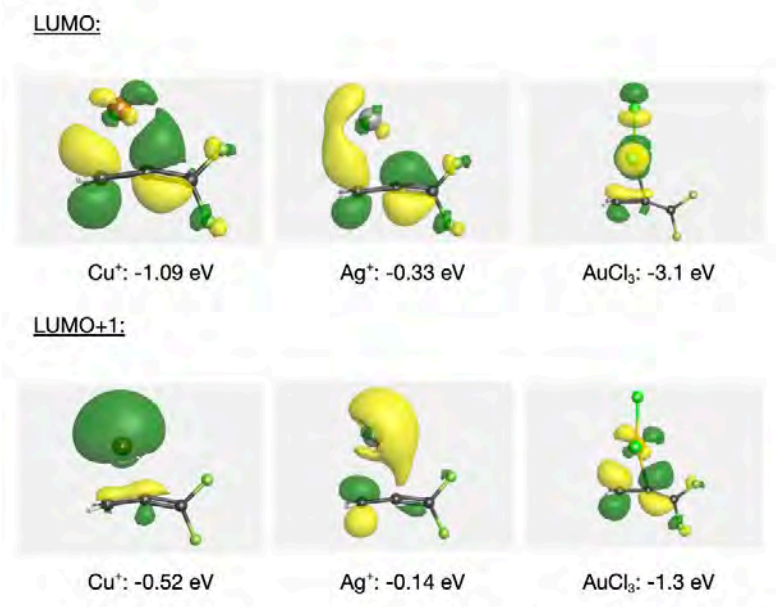


Figure 12. *LUMO and LUMO+1 orbital shapes and energies for Cu^+, Ag^+, and $AuCl_3$ coordinating on the hydrogen side of gem-difluoroallene.*

The LUMO for Cu^+ and Ag^+ complexes, along with the LUMO+1 of $AuCl_3$, share similar shapes, and it shows why γ-addition is feasible under orbital-controlled conditions. Additionally, the LUMO and LUMO+1 orbitals are not concentrated on the α carbon for all three complexes. This suggests that α-addition must be governed by favorable electrostatic interactions.

5. Conclusion

Through DFT analysis, the mechanism of AgF nucleophilic addition to *gem*-difluoroallene became clearer. It first involves the coordination of Ag^+ on the non-fluorinated side of a *gem*-difluoroallene moiety, which is lower in energy than coordination on the fluorinated side. Contrary to some previous mechanistic explanations, this intermediate would not involve a carbocation on the α carbon, which would explain why ^{19}F spectroscopy had failed to find such a structure. A subsequent attack of an F^- nucleophile on the α carbon produces the final product. While the fluorine anion could attack the γ position, this process is kinetically and thermodynamically less favorable than α-addition.

It has been noted in previous experiments that this nucleophilic α-addition tends to form products with E isomerism if one of the hydrogens is substituted by a larger group. With the new bonding structure of Ag^+ to *gem*-difluoroallene, substituted groups would form a dihedral angle of 90° with the fluorine atoms, and the allene structure only bends slightly. This means that the strain between the substituted group and the CF_2 group is minimal. Rather, it is the attacking F^- nucleophile that undergoes unfavorable steric interactions with the substituted group when approaching from the Z face to the intermediate structure with Ag^+ bonded to the hydrogen-side, which is why attacking on the E face is less energetically demanding.

While the focus of this study was AgF and nucleophilic α-addition, the metal coordination step and explanation of E-stereoselectivity applies to copper and gold mediated nucleophilic reactions with *gem*-difluoroallene. All three metals form a bonding interaction between its d_{z^2} orbital and the HOMO-1 orbital of *gem*-difluoroallene alongside a weaker back-bonding interaction with another one of its d-orbital. Analyses of the LUMO and LUMO+1 shapes for copper, silver, and gold intermediates show that γ-addition on *gem*-difluoroallene is possible under orbital-controlled conditions while α-addition relies on electrostatic interactions.

Hopefully, these findings provided insights into how metal-mediated, nucleophilic addition to *gem*-difluoroallene functions. Additionally, the methods employed in this study can be utilized to investigate more complicated reactions, explain experimental findings, and develop new reaction pathways involving this unique structural moiety.

References

(1) O'Hagan, D. *Chem. Soc. Rev.* **2008**, *37* (2), 308–319.
(2) Koley, S.; Altman, R. A. *Israel Journal of Chemistry* **2020**, *60* (3-4), 313–339.
(3) Xiong, Y.; Zhang, X.; Huang, T.; Cao, S. *The Journal of Organic Chemistry* **2014**, *79* (14), 6395–6402.
(4) Zhang, X.; Lin, Y.; Zhang, J.; Cao, S. *RSC Advances* **2015**, *5* (11), 7905–7908.
(5) Zhang, J.; Xu, C.; Wu, W.; Cao, S. *Chemistry - A European Journal* **2016**, *22* (29), 9902–9908.
(6) Ichikawa, J.; Wada, Y.; Okauchi, T.; Minami, T. *Chemical Communications* **1997**, No. 16, 1537–1538.
(7) Ichikawa, J.; Sakoda, K.; Wada, Y. *Chemistry Letters* **2002**, *31* (3), 282–283.
(8) Ichikawa, J.; Fujita, T.; Sakoda, K.; Ikeda, M.; Hattori, M. *Synlett* **2012**, *24* (01), 57–60.
(9) Li, J.; Rao, W.; Wang, S.-Y.; Ji, S.-J. *The Journal of Organic Chemistry* **2019**, *84* (18), 11542–11552.
(10) Lu, X.; Wang, Y.; Zhang, B.; Pi, J.-J.; Wang, X.-X.; Gong, T.-J.; Xiao, B.; Fu, Y. *Journal of the American Chemical Society* **2017**, *139* (36), 12632–12637.
(11) Ohashi, M.; Kambara, T.; Hatanaka, T.; Saijo, H.; Doi, R.; Ogoshi, S. *Journal of the American Chemical Society* **2011**, *133* (10), 3256–3259.
(12) Ohashi, M.; Saijo, H.; Shibata, M.; Ogoshi, S. *European Journal of Organic Chemistry* **2012**, *2013* (3), 443–447.
(13) Dai, W.; Xiao, J.; Jin, G.; Wu, J.; Cao, S. *The Journal of Organic Chemistry* **2014**, *79* (21), 10537–10546.
(14) Dai, W.; Wu, W.; Cao, S. *Organic Letters* **2015**, *17* (11), 2708–2711.
(15) Lu, C.-J.; Yu, X.; Chen, Y.-T.; Song, Q.-B.; Wang, H. *Organic Chemistry Frontiers* **2020**, *7* (16), 2313–2318.
(16) Ichikawa, J.; Yokota, M.; Kudo, T.; Umezaki, S. *Angewandte Chemie International Edition* **2008**, *47* (26), 4870–4873.
(17) Yokota, M.; Fujita, D.; Ichikawa, J. *Organic Letters* **2007**, *9* (22), 4639–4642.
(18) Shen, Q.; Hammond, G. B. *Journal of the American Chemical Society* **2002**, *124* (23), 6534–6535.
(19) Dolbier, W. R.; Burkholder, C. R.; Winchester, W. R. *The Journal of Organic Chemistry* **1984**, *49* (9), 1518–1522.
(20) Dolbier, W. R.; Burkholder, C. R.; Piedrahita, C. A. *Journal of Fluorine Chemistry* **1982**, *20* (5), 637–647.
(21) Dolbier, W. R.; Burkholder, C. R. *The Journal of Organic Chemistry* **1984**, *49* (13), 2381–2386.
(22) Kühnel, M. F.; Lentz, D. *Dalton Transactions* **2009**, No. 24, 4747.
(23) Mae, M.; Hong, J. A.; Xu, B.; Hammond, G. B. *Organic Letters* **2006**, *8* (3), 479–482.
(24) Fuchibe, K.; Ueda, M.; Yokota, M.; Ichikawa, J. *Chemistry Letters* **2012**, *41* (12), 1619–1621.
(25) Fuchibe, K.; Abe, M.; Sasaki, M.; Ichikawa, J. *Journal of Fluorine Chemistry* **2020**, *232*, 109452.
(26) Luo, H.; Zhao, Y.; Wang, D.; Wang, M.; Shi, Z. *Green Synthesis and Catalysis* **2020**, *1* (2), 134–142

(27) Gao, B.; Zhao, Y.; Ni, C.; Hu, J. *Organic Letters* **2013**, *16* (1), 102–105.
(28) Liu, C.; Zhu, C.; Cai, Y.; Yang, Z.; Zeng, H.; Chen, F.; Jiang, H. *Chemistry – A European Journal* **2020**, *26* (9), 1953–1957.
(29) Qi, S.; Gao, S.; Xie, X.; Yang, J.; Zhang, J. *Organic Letters* **2020**, *22* (13), 5229–5234.
(30) Breton, G. W. *Journal of Physical Organic Chemistry* **2021**, *35* (3).
(31) Shan, C. C.; Dai, K. Y.; Zhao, M.; Xu, Y. H. *European Journal of Organic Chemistry* **2021**, *2021* (29), 4054–4058.

ViT4SF: Vision Transformers for Solar Forecasting

Pranav Virupaksha

Author Background: *Pranav Virupaksha grew up in the United States and currently attends Lynbrook High School in San Jose, California in the United States. His Pioneer research concentration was in the field of computer science and titled "Computers That See: Exploring Computer Vision."*

Abstract

A major obstacle to the integration of solar panels into electricity grids around the world is the problem of high variability in terms of power output. Power output from solar panels can be affected by various constantly changing factors, including cloud cover and the sun's changing position. To effectively anticipate and manage power generation and consumption, electricity grid operators must be able to estimate the power output from their various energy generation methods, including solar panels. Thus, the problem of solar forecasting, or estimating future solar panel power output, has been the subject of extensive research, with varying methods and types of inputs, including sequences of past power outputs (power logs) and ground-based sky images. This paper presents and evaluates a promising method for further research in the solar forecasting task: ViT4SF (Vision Transformers for Solar Forecasting). After training our relatively simple yet powerful approach on the newly released SKy Images and Photovoltaic Power Generation Dataset (SKIPP'D) of thousands of sky images and corresponding power outputs, we show promising results and improvements to the SKIPP'D benchmark results. We furthermore experiment with two variants of ViT4SF: ViT4SF Base, with only sky images as input, and ViT4SF w/Power Log, with both sky images and power logs as input, and evaluate their performances under various weather conditions.

1. Introduction

Solar panels are quickly becoming a large part of the world's power supply. A major obstacle to the integration of solar panels into electricity grids around the world, however, is their variability in terms of their power output. This fluctuation can be caused by multiple dynamic factors, including cloud cover

and the changing position of the sun, which make it difficult for electricity grid operators to effectively forecast the expected output from their solar sources of energy. Electricity grid operators need these forecasts to manage power generation and consumption, as well as to set financially sound prices for power.

To this end, the task of solar forecasting has been explored extensively in recent years, marked by the use of various machine learning algorithms to estimate the power output of a solar panel. Multiple approaches and data inputs have been explored in the past, including sequences of previous power outputs (Konstantinou et al., 2021), sequences of previous ground-based images of the sky above the panel (Zhao et al., 2019), and a combination of these two types of data (Sun et al., 2019). These algorithms are all aimed at the target of solar forecasting: predicting a numerical solar power output value at a certain forecast horizon (e.g. 15 minutes ahead).

This work contributes to the branch of sky image-based solar forecasting algorithms by proposing and presenting an evaluation of a novel method of using vision transformers on sequences of sky images to predict solar power output. We furthermore test the effects of adding past power outputs as a secondary input to the network, creating a base and modified model. Both models are trained on the newly released SKy Images and Photovoltaic Power Generation Dataset (SKIPP'D) by Nie et al. (2022) and are evaluated in comparison to the associated benchmark results for the dataset. To keep metrics comparable to this benchmark, we choose to train our networks to forecast power outputs 15 minutes ahead. While relatively simple in its approach, our method proves itself to be a capable and promising direction for further research in solar forecasting algorithms.

2. Background

In this section, we will introduce background concepts needed to understand vision transformers, including recurrent neural networks, long short-term memory, convolutional neural networks, and transformers.

2.1. Recurrent Neural Networks (RNNs)

RNNs are a class of neural networks that are adapted to sequential data (e.g. text, audio, or videos). These networks use information from previous items in the input sequence to inform their outputs.

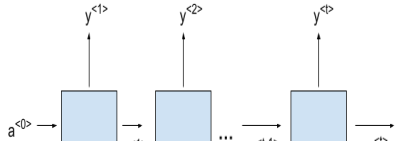


Figure 1. *RNN Architecture*

As seen in Figure 1, an RNN takes in two inputs: $a^{<t-1>}$ (the activation of the previous timestep) and $x^{<t>}$ (the current x-value) to generate $y^{<t>}$ (the prediction for the current time step), and $a^{<t>}$ (the activation for the current timestep). The two inputs are concatenated and passed through a fully-connected layer to yield the first output $a^{<t>}$. This is then passed into another fully-connected layer to yield the secondary output $y^{<t>}$. The outputted activation $a^{<t>}$ is then passed on to the timestep *t+1*, with $x^{<t+1>}$ as input, and this process continues until all of the items in the input sequence *x* are passed into the network.

The RNN can be described with the following equations, where g_1 and g_2 symbolize the fully-connected layers, W_{aa}, W_{ax}, and W_{ya} symbolize their associated weights, and b_a and b_y symbolize their associated biases:

$$a^{<t>} = g_1(W_{aa}a^{<t-1>} + W_{ax}x^{<t>} + b_a)$$
$$y^{<t>} = g_2(W_{ya}a^{<t>} + b_y)$$

RNNs have proved themselves capable in the tasks of text generation, text classification, and many more. However, these models have some notable disadvantages, including that their forward passes can be slow because of their iterative nature, and that they do not perform well on longer sequences because the information from earlier in the sequence slowly fades from the activation being passed between timesteps. It is also difficult to train RNNs on longer sequences because the gradients of timesteps can either slowly degrade or be amplified as one goes backward in time. These are known as the vanishing or exploding gradient problems.

2.2. Long Short-Term Memory (LSTM)

The LSTM introduces several additions that help solve the vanishing and exploding gradient problems. It introduces the idea of a cell state which stores the long-term memory of the network, with several “gates” with varying functions to help the network make sense of the data. As seen in Figure 2, the LSTM takes the activation and the cell state of the previous timestep and the x-value for the current timestep as inputs. It then outputs a cell state, a hidden state, and a y-value as its prediction for the timestep.

At each timestep, the previous activation $a^{<t-1>}$ and the current input $x^{<t>}$ are concatenated into a single vector which is fed into the various gates of the LSTM.

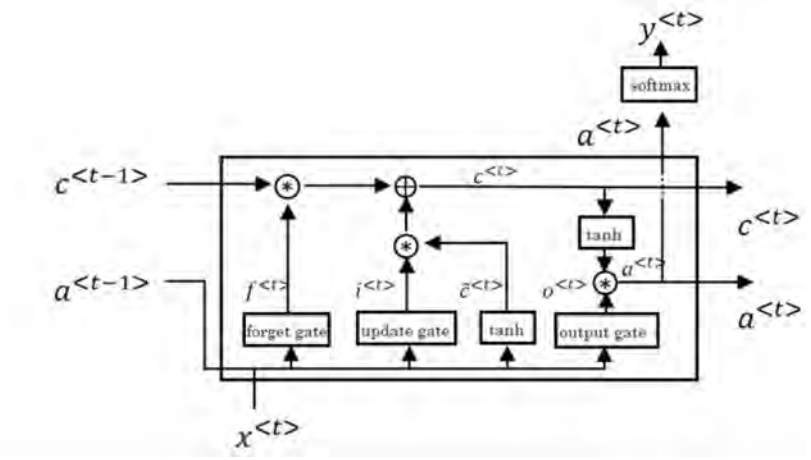


Figure 2: *Diagram of LSTM Architecture (Ng, 2017)*

2.2.1. Forget Gate

The forget gate's purpose is to determine which parts of the cell state to keep or discard. The concatenated vector composed of $a^{<t-1>}$ and $x^{<t>}$ is run through the forget gate, a fully connected layer with the Sigmoid activation function. Because the Sigmoid activation function is used, the output of the forget gate is restricted to values between 0 and 1. The resulting vector from the forget gate, $f^{<t>}$, is then elementwise multiplied with the cell-state $c^{<t-1>}$. Values in $f^{<t>}$ closer to 0 will suppress unwanted features in the cell state, while values closer to 1 will not have much of an effect.

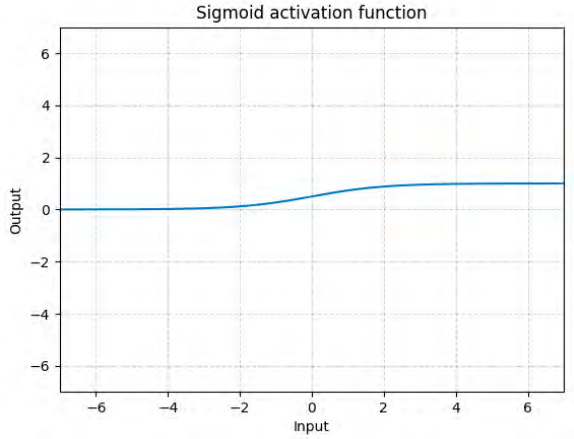


Figure 3: *Sigmoid Activation Function. Values are restricted between 0 and 1. (PyTorch, n.d.)*

2.2.2. Update Gate

The update gate's purpose is to generate a vector from the previous activation $a^{<t-1>}$ and current input $x^{<t>}$ to add new information to the cell state. The concatenated vector composed of $a^{<t-1>}$ and $x^{<t>}$ is run through both a fully-connected layer with the Tanh activation function and the update gate, which is a fully connected layer with the Sigmoid activation function. The results, $i^{<t>}$ and $\tilde{c}^{<t>}$, are then elementwise multiplied to yield a vector with new information for the cell state. This vector is then combined with $c^{<t-1>}$ using elementwise addition to yield the cell state for the current timestep, $c^{<t>}$.

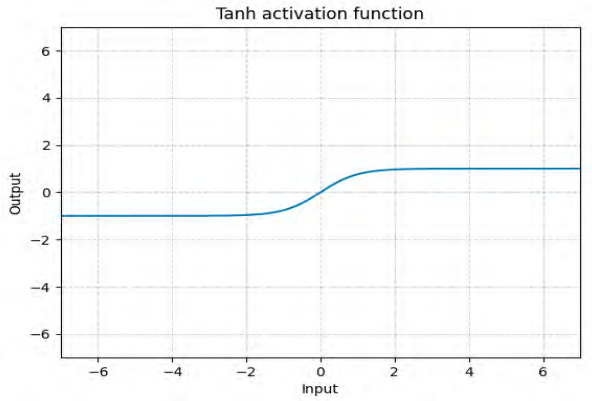


Figure 4: *Tanh Activation Function. Values are restricted between -1 and 1. (PyTorch, n.d.)*

2.2.3. Output Gate

The output gate's purpose is to use information from the current cell state $c^{<t>}$, the previous activation $a^{<t-1>}$, and the current input $x^{<t>}$ to generate the current activation $a^{<t>}$. The concatenated vector composed of $a^{<t-1>}$ and $x^{<t>}$ is passed through the output gate, which is a fully-connected layer with the Sigmoid activation function. Simultaneously, the current cell state $c^{<t>}$ is passed through the Tanh activation function. The resulting vectors are then elementwise multiplied to yield the activation for the current timestep, $a^{<t>}$.

The activation $a^{<t>}$ can then be run through outside fully connected layers to yield the prediction for the timestep, $y^{<t>}$.

Using equations, the operations of the LSTM can be described as the following:

$$i_t = \sigma(x_t U^i + a_{t-1} W^i + b_i)$$
$$f_t = \sigma(x_t U^f + a_{t-1} W^f + b_f)$$
$$o_t = \sigma(x_t U^o + a_{t-1} W^o + b_o)$$
$$\tilde{C}_t = tanh(x_t U^g + a_{t-1} W^g + b_g)$$
$$C_t = \sigma(f_t * C_{t-1} + i_t * \tilde{C}_t)$$
$$a_t = tanh(C_t) * o_t$$

2.3. Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a powerful class of neural networks that have proven to be extremely capable with image data. CNNs are based on the central operation of convolutions, which allow the network to work well with image data. CNNs are commonly applied in the tasks of image classification, object detection, semantic segmentation and many more. A common CNN architecture, in this case for the image classification task, can be seen in Figure 5.

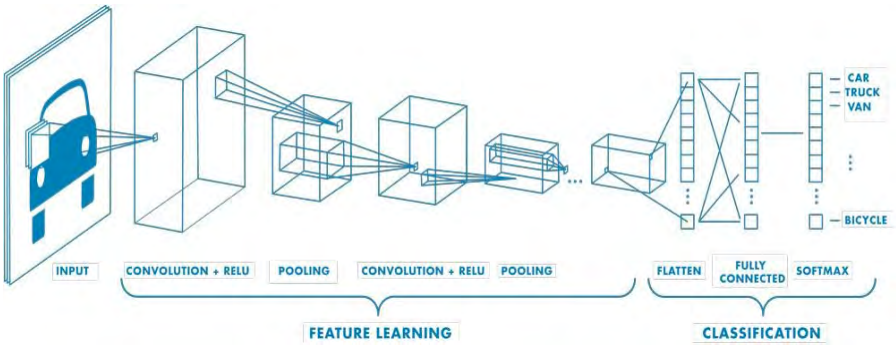


Figure 5: *Common Convolutional Neural Network architecture for image classification. (MathWorks, n.d.)*

2.3.1. Convolutional Layer

The CNN is centered around the use of a Convolutional Layer, which performs the convolution operation on images. The Convolutional Layer applies a learned matrix, or kernel, with a defined size across an image. This kernel, and the respective area that it slides over, are dot product multiplied to yield a feature map that summarizes the high-level features of the images. Each neuron in a convolutional layer consists of a learnable kernel that is applied to the image, with the goal being to extract useful features in an image that may help the network with its task. A visualization of the convolutional operation can be seen in Figure 6.

As seen in Figure 6, as the kernel slides over the image, it passes over equally sized patches. The kernel and these patches are elementwise dot product multiplied and the result of this calculation takes its place on the resulting feature map. In the case below, the kernel applied is a 2D matrix and has one channel because the image is also 2D and has one channel. However, if the image has multiple channels, as is seen in RGB images, the kernel would have an equivalent depth as well.

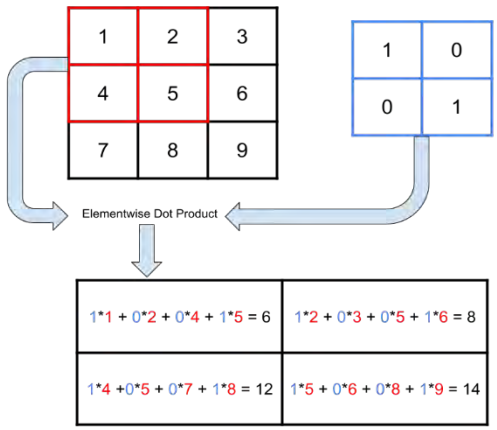


Figure 6: *Illustration of the convolution operation. The kernel, shown in blue, slides over and interacts with equally-sized patches of the image, shown in red.*

Various activation functions can be applied to the resulting feature maps, including the Tanh and Sigmoid functions, which have been shown before, and the ReLU function. As seen in the graph below, the ReLU function suppresses negative values while preserving positive values.

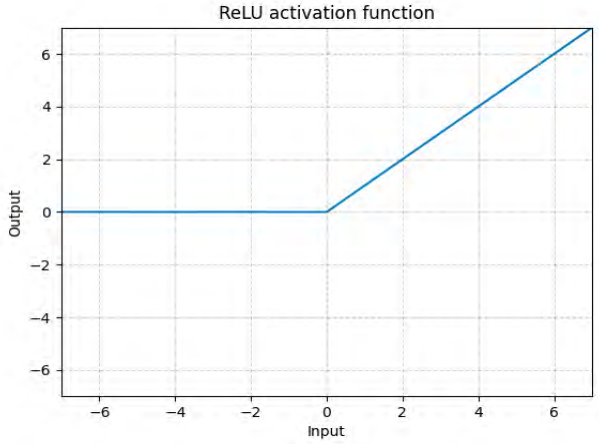


Figure 7: *ReLU Activation Function (PyTorch, n.d.)*

2.3.2. Pooling Layer

The pooling layer downsamples images and feature maps and helps reduce their spatial sizes. While there are multiple types of pooling layers, the most relevant in the case of this paper is the maximum pooling layer. In maximum pooling, the layer applies a kernel across the image that only returns the largest value to the feature map. As seen in Figure 8, the maximum pooling

layer returns only the maximum value of each patch of the image it passes through.

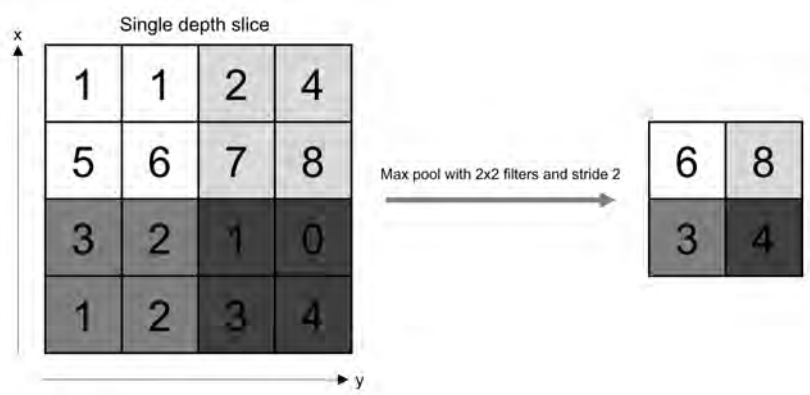


Figure 8: *Image of the Pooling Layer Operation by (Bonaccorso et al., 2018)*

Once important features are extracted from the image using convolutional and pooling layers, they are flattened into a 1-dimensional vector to be fed into a fully connected network. This fully connected network can be configured to accommodate multiple tasks including classification and regression tasks, such as image classification and object detection.

2.4. Transformers

2.4.1. Seq2Seq Models

Composed of LSTMs or RNNs, Seq2Seq models take in a sequence as input and generate a sequence as output. Applications of these models include machine translation and sentence completion. As seen in Figure 9, Seq2Seq models consist of two parts: an encoder, and a decoder, both being composed of LSTMs or RNNs.

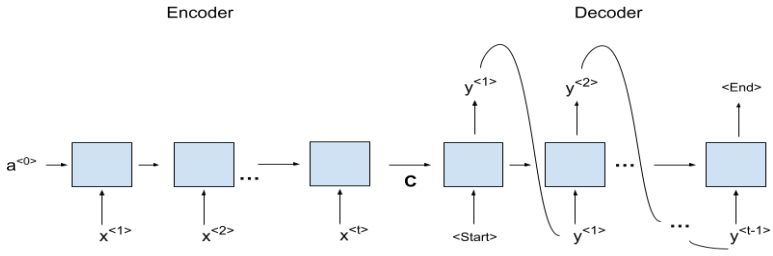


Figure 9: *Seq2Seq Model Architecture*

The encoder runs through the input sequence and passes its final activation, also known as the context vector, to the decoder. The decoder, with

the combined input of a <Start> token and the context vector C, generates an output $y^{<1>}$ and an activation. Both of these outputs are then passed through the decoder until it generates an <End> token. As a result, the decoder generates a sequence based on the encoded input sequence.

2.4.2. Transformer Architecture

While Seq2Seq models are extremely capable and can be applied to many tasks, they carry over the problems of LSTMs and RNNs. Because inputs need to be fed into the networks sequentially, the training process for these networks is slow. This is because one cannot properly take advantage of the parallel processing of modern machine learning computing hardware, such as GPUs. To take advantage of such parallelization, the Transformer architecture was introduced by Vaswani et al (2017).

As seen in Figure 10, like the Seq2Seq architecture, the Transformer consists of encoder and decoder components. In the case of the Transformer, identical encoders are stacked on top of each other in the encoding block, and decoders are stacked in the decoding block. Notably, the output from the encoding block is passed into every single decoder in the decoding block.

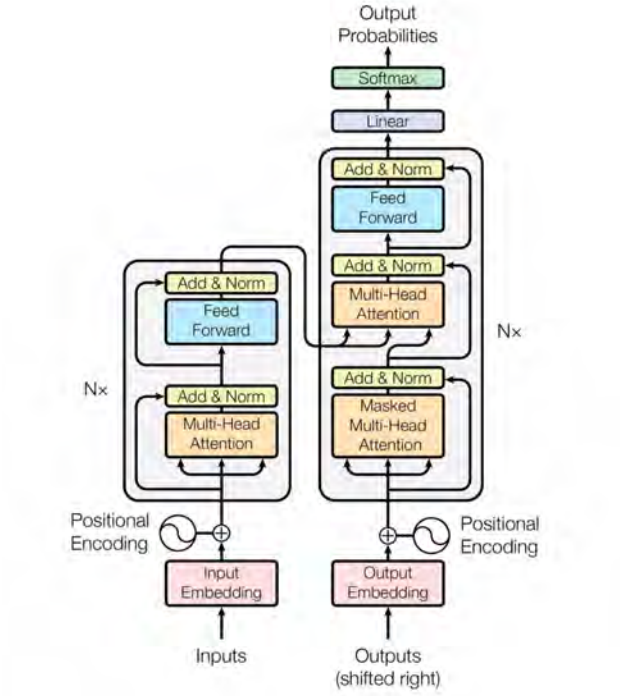


Figure 10: *Transformer Architecture showing the encoding and decoding components. (Vaswani et al., 2017)*

The input to the Transformer, like Seq2Seq models, consists of a sequence, such as a sentence of words. This sequence is first embedded into a sequence of vectors that the Transformer can understand, and then information about the positions of items in the sequence is added. This “positional encoding”

is created using a simple sine or cosine function-based formula and has the same dimension as an embedded vector. The result of this step is a sequence of vectors consisting of the sum between the embedded input vectors and their corresponding positional encodings. This is then fed into the encoder.

First, these inputs are passed through a self-attention module. At a high level, self-attention helps the network look at other items in the sequence for context as the network encodes an input item. For example, if the input sequence were made up of the words “I bought the car and it was fast”, self-attention would allow the network to pay attention to context from the words “car” and “fast” when encoding the word “it”.

The outputs and inputs of the self-attention module are then summed and passed to a fully connected layer. Again, the outputs and inputs of the fully connected layer are summed together, yielding the output of the encoder. The output of this encoder can then be passed to further encoders stacked on top of it. If the encoder is at the top of the encoder stack, its output is passed to each decoder in the decoder stack, as an input into the decoder’s self-attention module (Figure 11).

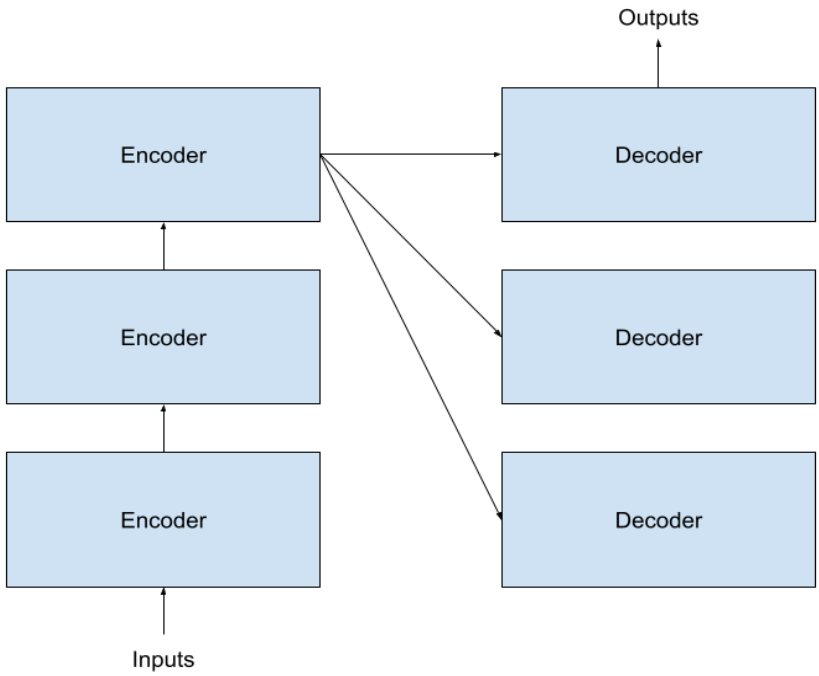


Figure 11: *Demonstrating how encoder outputs are passed on to decoders. The encoder at the top of the encoder stack passes its outputs to each of the decoders in the decoder stack.*

The decoder component of the Transformer architecture shares many characteristics with the encoder component, with a few differences. First, similarly to decoders in Seq2Seq models, the inputs for the decoder are “shifted

right," essentially meaning that they consist of the outputs of previous timesteps, using a specific token for the first input. Thus, the decoder can only use self-attention to use outputs from previous timesteps for context. The decoder uses a masked self-attention module for this purpose, ensuring that the decoder outputs can only depend on previously generated outputs. After being generated from the decoder, output vectors are passed into a fully connected layer which is configured based on the application of the Transformer network, such as text generation.

The setup of the Transformer architecture allows for inputs to be passed into the network in parallel, thus allowing for faster training times with hardware such as GPUs.

2.4.3. Vision Transformers

Proposed by Dosovitskiy et al. (2021), Vision Transformers (ViTs) repurpose the original transformer architecture for the computer vision field and image classification task by making a few notable changes related to the input of image data. To harness the power of the transformer architecture on images, ViTs must first convert an image into a sequence. This is performed by dividing the image into patches, flattening these patches into vectors, and embedding them using a fully connected layer. Positional encodings are then added to these vectors, with an additional learnable embedding being added to the beginning of the sequence of vectors. A visualization of this is shown in Figure 12.

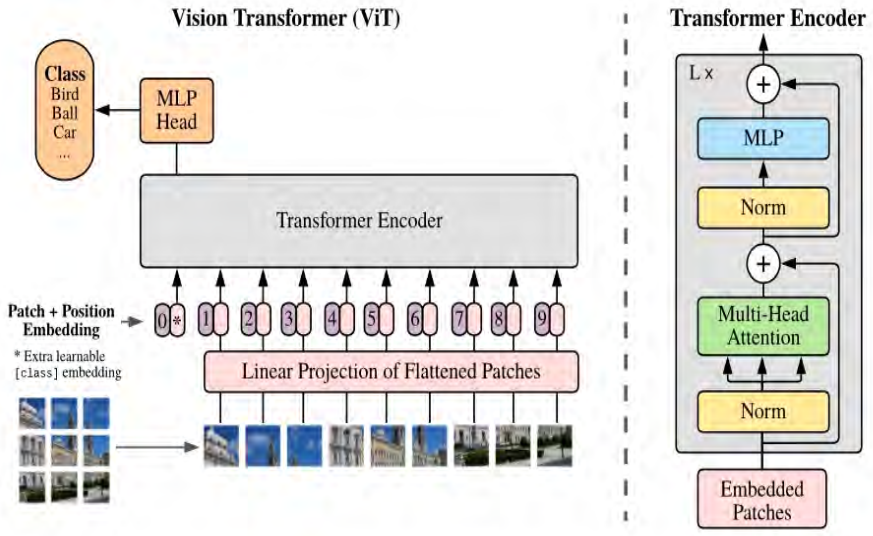


Figure 12: *ViT Architecture. "MLP Head" refers to a block of fully connected layers that make the image classification prediction. (Dosovitskiy et al., 2021)*

Notably, ViTs only use the encoder part of the transformer architecture, connecting the encoder output to a block of fully connected layers which makes the image classification predictions. Using this approach, Vision Transformers

have proven themselves to be extremely capable at feature extraction and image classification, achieving comparable results to CNNs on various benchmarks.

3. Related Works

In this section, we will provide a short overview of the area of solar forecasting research, focusing on the more relevant methods to this paper while briefly mentioning others.

According to Sobri et al. (2018), the vastly extensive area of solar forecasting research can be categorized into three main categories: time series statistical methods, physical methods, and ensemble methods. For the sake of brevity, we will focus on the more relevant category of physical methods here, which are based on data captured from the physical world, consisting of numerical weather prediction, sky imagery, and satellite imaging. We will also address some specific works that are relevant to our work within the time-series category.

In addition to these areas, it is also worth noting that various forecast horizons are researched in the solar forecasting task, ranging from long-term forecasts of power outputs years ahead, to medium-term monthly forecasts, to short-term hourly or minute-based forecasts. In this paper, we will be presenting our work in the short-term forecast horizon area and will thus be mainly focusing on short-term solar forecasting algorithms.

3.1. Numerical Weather Prediction

Numerical weather prediction (NWP) methods focus on the processing of physical data, such as incoming solar radiation and topography, by mathematical models to predict future weather variables, such as temperature. As seen in Fernandez-Jimenez et al. (2012), NWP models can be combined with machine learning methods and applied to the task of solar forecasting. In this paper, the authors use a three-module structure consisting of a sequence of two NWP models feeding into a machine learning model. Various machine learning methods such as k-NN (k-Nearest Neighbors) and ANN (Artificial Neural Network) are tested for this model, with the ANN performing the best.

3.2. Sky Imagery

Various methods using sky imagery as input have been proposed for the task of solar forecasting. Sky image data is collected by ground-based, often hemispherical cameras, sometimes known as sky imagers. According to Sobri et al. (2018), sky images are beneficial because they provide valuable visual information on quick-changing cloud cover that can reduce the power output of solar panels. Zhao et al. (2019) introduce a 3D-CNN-based model taking multiple consecutive sky images as input for effective feature extraction for the solar forecasting task. In 3D-CNNs, a three-dimensional kernel is used on a stack of images to extract both spatial and temporal information. With this, Zhao et al. aim to extract information about the motion of cloud cover to further inform predictions. Conversely, Feng and Zhang (2020) introduce their model SolarNet which takes in only one sky image as input, showing that this less

computationally expensive approach is somewhat viable for the task. Other works propose hybrid architectures using a combination of sky images and weather variables, as seen in Alani et al. (2021).

3.3. Satellite Imagery

Satellite imagery-based methods use visual and infrared sensors to gain information about cloud patterns from space. Like NWP methods, these features can be combined with machine learning methods to yield power output forecasts. For example, as seen in Marquez et al. (2013), cloud position and motion data are extracted from satellite images and then inputted into an ANN to yield a forecast.

3.4. LSTM-Based Methods

A variety of works have explored the application of LSTMs to the task of solar forecasting, proposing and evaluating different types of input and forecast horizons. Lee and Kim (2019) input meteorological data such as temperature, humidity, and cloudiness, as well as the date, into an LSTM to predict outputs for a 1-hour forecast horizon. Following a similar line of thought, Gao et al. (2019) use meteorological information based on NWP such as air temperature and level of cloudiness for a 1-hour forecast horizon. In contrast, Konstantinou et al. (2021) present a stacked LSTM network that uses only previous power outputs and performs time-series forecasting to predict future power outputs of a forecast horizon of 1.5 hours.

3.5. SKIPP'D Dataset and Baseline Model

Because of the varying metrics and forecast horizons used in solar forecasting research, it is difficult to compare methods. We thus choose to use a dataset and baseline to test the effectiveness of our model, more specifically the SKIPP'D dataset and its accompanying baseline model.

The SKIPP'D dataset, created and introduced by Nie et al. (2022), is a sky images and solar power output dataset geared toward two tasks: solar power forecasting, the focus of this paper, and solar power nowcasting, which focuses on predicting solar power output as it is being generated. The dataset consists of around 380,000 total sky image and solar power output pairs, which can be further processed to create a subset of samples for the forecasting task. This will be further expanded upon in the next section.

In terms of its data collection parameters, the dataset was collected on top of buildings on the Stanford University campus on specific days from the years 2017 to 2019. Furthermore, samples in this dataset were collected at one-minute intervals, enabling much finer forecast horizons to be explored. The temporal resolution aspect of the dataset is one of its distinguishing features: other solar forecasting sky image datasets, such as those introduced by Stoffel and Andreas (2015) and Augustine et al. (2000), have comparatively lower temporal resolutions. While the dataset introduced by Pedro et al. (2019) has

one-minute intervals as well, the overall size of the dataset is small compared to the SKIPP'D dataset.

The sky images are collected via a 6-megapixel, 360-degree fish-eye camera that is placed 125 meters away from the solar panel from which the power outputs are recorded. The camera records videos of resolution 2048 x 2048 which are processed down to 64 x 64 images. The solar panel, with a 30.1 kW capacity, has an elevation angle of 22.5 degrees, and an azimuth of 195 degrees.

The baseline model, known as SUNSET (Stanford University Neural Network for Solar Electricity Trend), is introduced by Sun et al. (2019) and consists of a convolutional neural network, with an accompanying series of fully connected layers used to make the prediction. The model takes in two inputs: one is a sequence of images from the past 15 minutes, while another is a sequence of past power outputs, better known as a power log. To input the sequence of images, which are in a tensor of shape 16x64x64x3, the authors simply reshape the tensor into the shape 64x64x48, seemingly not regarding the sequential nature of the images. Once feature vectors have been extracted from this data, they are concatenated with the power log of 16 values. The resulting concatenated vector then passes through the fully connected layers at the end of the network to yield the solar power output forecast.

Because the authors of the SUNSET baseline do not take advantage of the sequential nature of the past images, we choose to explore how this might be leveraged with vision transformers, which, as explained above, have been shown to work well with sequential data.

4. ViT4SF

4.1. Data Preparation:

As introduced in the prior section, the SKIPP'D Dataset is used for the training of our models. In particular, we use the forecasting subset of the original dataset. Raw images and their corresponding power output values, along with the code to create the forecasting dataset from this unrefined data, are made publicly available[1] by the dataset authors on GitHub. Using this code, the forecasting dataset was replicated to the best of our ability, with the same conditions stated in the original paper.

[1] Code is available at https://github.com/yuhao-nie/Stanford-solar-forecasting-dataset.

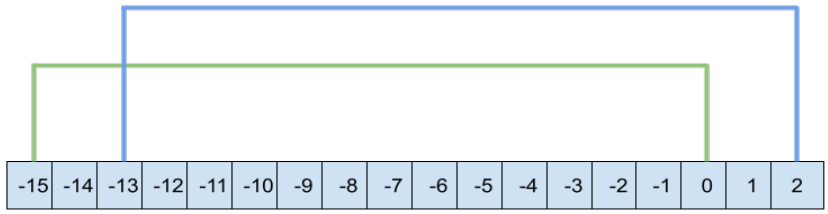


Figure 13: *Boxes represent example timestamps and their respective values in the dataset, with numbers x corresponding to t = x. The connecting lines show the first and last timestamps for valid, consecutive forecasting samples.*

As shown in Figure 13, inputs taken from the forecasting dataset consist of stacks of sky images and their corresponding power outputs collected from t = -15 minutes to t = 0 minutes. Because the dataset has a time resolution of one minute, this results in the input being a stack of 16 images. The target consists of the solar power output at t = 15 minutes. Furthermore, samples are taken at two-minute intervals: if a sample were centered around $t=t_i$, the next valid sample would be centered around $t=t_{i+2}$.

Using the authors' code and the unrefined public dataset, we generate 130,412 forecasting samples for the training and validation datasets from an original 349,372 images and their respective outputs. We then shuffle the samples and use a training/validation split of 0.25.

We do not use augmentations on the image data for the sake of preserving hypothesized salient features in the data. We hypothesize that the brightness of the image, as well as the positioning of the sun in the image, may play a vital role in the prediction of solar power output. Thus, we do not apply any geometric, brightness, or color-changing augmentations to the images.

Because we use a vision transformer for the encoding of the sky images, we need to use single images as input. Thus, we unroll and stitch the stack of 16 - 64x64 images into one 256 x 256 image, following the positioning of the images in the stack. To keep the sequential nature of the image intact, the ViT we use is modified to embed 64 x 64 patches with a stride of 64, meaning that each patch that is inputted into the transformer will be a singular sky image.

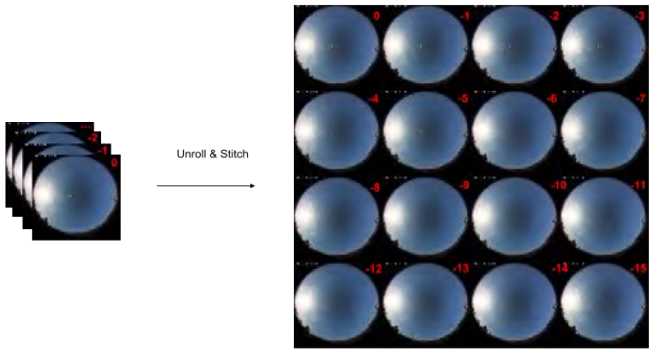


Figure 14: *Visualizing the unrolling and stitching of the stack of images. A number x on an image marks t=x in the sequence.*

4.2. Architecture Overview

The base architecture of ViT4SF uses a modified Vision Transformer as a feature extractor, with the encoding from this step being passed to a fully connected layer which makes the actual solar output regression prediction. We do not use any activation functions on this fully connected layer.

Much like the SUNSET model, the ViT4SF w/Power Log variant concatenates a one-dimensional vector consisting of the power log to the encoded vector from the ViT. The resulting vector of length 784 is then inputted into a series of fully connected layers, each using the ReLU activation function. The final fully connected layer, as with the base ViT4SF model, does not use any activation functions.

In both cases, we modify the ViT to embed 64x64 patches with a stride of 64, with the goal being to input the original 64x64 images for each timestamp separately and sequentially.

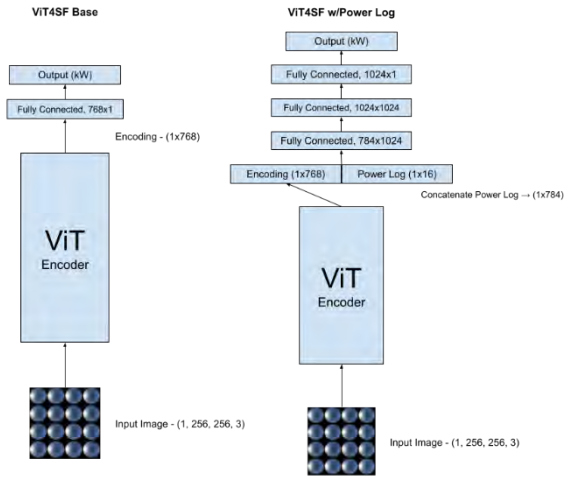


Figure 15: *Model architectures for ViT4SF Base and its w/Power Log variant*

4.3. Implementation

We use the popular Transformers[2] library from HuggingFace (Wolf et al., 2019) for the initialization of the ViT, and the machine learning framework PyTorch[3] (Paszke et al., 2019) for further setup and training. The default HuggingFace Vision Transformer configuration is used, except for the patch size and stride being changed to 64, and the final "classification" layer of the network being removed and changed to randomly initialized fully connected layers defined by PyTorch.

We train the base model for 150 epochs using the SGD optimizer, with a learning rate of 1e-6 and a momentum of 0.9. We use a batch size of 32 for the training and validation of the network and mean squared error (MSE) loss. All training operations are performed through PyTorch.

For the variant model using power logs, the only differences in our training process are that we use the Adam optimizer with a learning rate of 5e-6 and train the network until the validation loss stops decreasing.

5. Analysis

For the creation of the testing dataset, we again use the SKIPP'D dataset authors' code with the slight modification of using a one-minute interval instead of a two-minute interval between collected samples: if a sample is centered around $t=t_i$, the next valid sample is centered around $t=t_{i+1}$. This is done to ensure that we have enough samples for a more complete evaluation of the model. For the test set, the authors also provide specific dates which can be used to obtain either sunny or cloudy images to gain more information on the generalizability of the model. We generate 5654 samples from sunny days and 5432 samples from cloudy days.

As is done in the SKIPP'D paper, we evaluate the models on root mean squared error (RMSE), mean absolute error (MAE), and forecast skill. We use the author-provided code to create the persistence model for the forecast skill metric. The persistence model is a naive, equations-based baseline that uses physical details relating to the angle of the solar panel, the time of day, and the date to predict the solar power output. This persistence model is then compared against our model to yield the forecast skill metric.

[2] https://huggingface.co/docs/transformers/index
[3] https://pytorch.org/

Sunny Image Cloudy Image

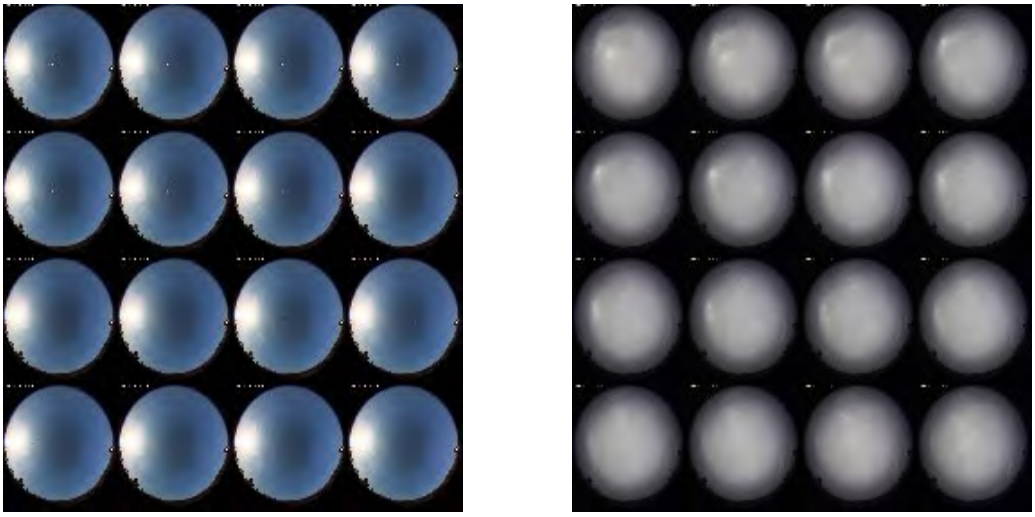


Figure 16: *Examples of Sunny and Cloudy images used for testing*

The formula for the forecast skill metric on a given sample is defined as:

$$\text{Forecast Skill} = 1 - (\text{RMSE}_{\text{model}} / \text{RMSE}_{\text{persistence}})$$

Table 1: *Comparison of results between models.*

Model	Test Set	RMSE (kW)	MAE (kW)	Forecast Skill (%)
SUNSET Forecast	Sunny Days	0.61	0.5	-45.8
	Cloudy Days	4.27	2.95	17.03
	Overall	3.03	1.71	16.44
ViT4SF Base	Sunny Days	**0.487**	0.397	**-13.43**
	Cloudy Days	5.05	3.42	1.37
	Overall	3.55	2.41	1.25
ViT4SF w/Power Log	Sunny Days	0.51	**0.384**	-18.55
	Cloudy Days	4.62	2.98	9.38
	Overall	3.27	2.1	9.14

As seen by the results in Table 1, our base model outperforms the SUNSET model on the sunny days test set in all categories. However, it underperforms the SUNSET model on the cloudy days test set. ViT4SF Base, like SUNSET, underperforms against the persistence model on sunny images while outperforming against the persistence model on cloudy images.

ViT4SF w/Power Log outperforms the SUNSET model and achieves comparable performance with the base model on sunny images. Furthermore, it

outperforms the base model on cloudy images, while achieving comparable results to the SUNSET model in the MAE metric. ViT4SF w/Power Log also underperforms against the persistence model on sunny images while outperforming against the persistence model on cloudy images.

Table 2: *Model performances with respect to MAPE and normalized MAPE (nMAPE) metrics.*

Model	Test Set	MAPE (%)	nMAPE (%)
ViT4SF Base	Sunny Days	**6.19**	**5.28**
	Cloudy Days	106.8	30.92
ViT4SF w/Power Log	Sunny Days	6.53	5.57
	Cloudy Days	**53.58**	**25.09**

We further evaluate our model on the Mean Absolute Percentage Error (MAPE) and normalized MAPE (nMAPE), another set of forecasting metrics. As seen in Table 2, the performances on these metrics reflect the performance observed on the metrics in Table 1. There is a marginal difference on sunny days, but a large disparity on cloudy days.

As seen in Figure 17, both models perform well on sunny data, with slight deviations being noted. However, on cloudy images, the models seem to suffer the most when there are sharp inclines or declines in the power being generated, which are caused by the variability of cloud cover. As demonstrated by the table results, the addition of power logs to the input for the model seems to make the predictions more stable and closer to the ground truth values.

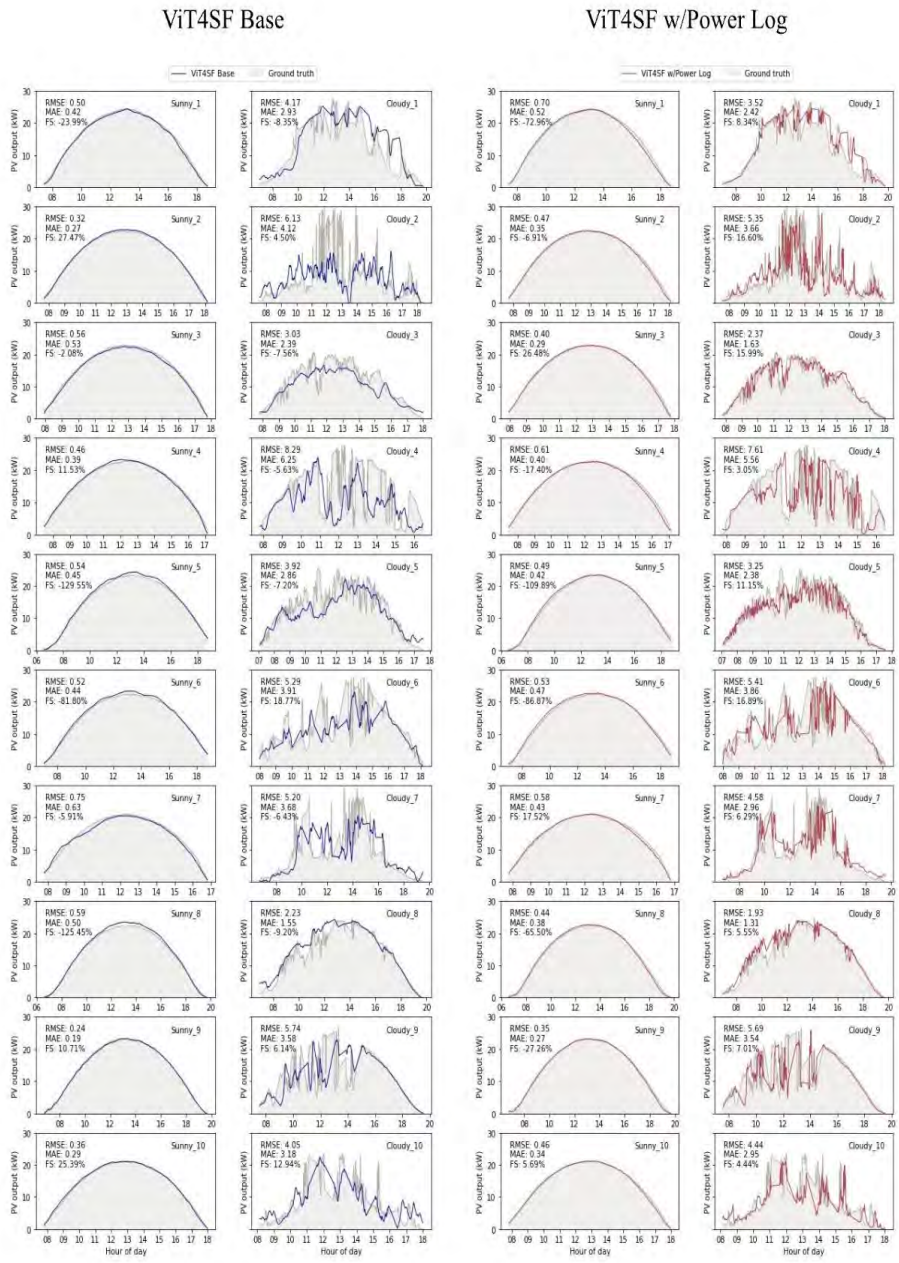


Figure 17: *Plots comparing the performance of both ViT4SF models to the ground truth. SKIPP'D author code was used to create this display.*

6. Future Work

This work provides a promising evaluation for exploring vision transformers in the area of solar forecasting research. It also provides interesting insights into the effectiveness of leveraging the sequential nature of solar forecasting input data which may be explored in the future. Because this approach uses vision transformers, which use the self-attention mechanism, future works can perform rigorous studies to determine the parts of the input image that the network pays attention to and are thus most useful. While we lightly experimented with this in the process of evaluating the network, we could not find any distinguishable patterns in this respect.

As discussed above, the SKIPP'D dataset provides the opportunity for researchers to work with not only sequences of past images, but also sequences of past power outputs. Thus, another possible direction for future research could evaluate a multimodal transformer that would take both images and past power outputs as input.

Furthermore, because of the sequential nature of the inputted images to the model, they can be essentially treated as video input. Future research could investigate the adaptation of video input-based machine learning algorithms, such as video transformers and 3D-CNNs, to this task.

Another important area to consider is the applicability of this model to conditions other than the ones it was trained on. The term *conditions*, in this case, includes:

1. The positioning of the solar panel and its capacity/other intrinsic properties
2. The geographic location of the solar panel
3. The positioning of the camera with respect to the solar panel and intrinsic camera properties.

Future works could explore the applicability of this work by fine-tuning and evaluating the model on other solar forecasting datasets using transfer learning.

7. Conclusions

In this work, we have presented and evaluated the use of vision transformers for the solar forecasting task, using images as the main input. We have furthermore evaluated the effects of adding power logs as an input to our proposed model and thus have created two models, ViT4SF Base and ViT4SF w/Power Log. We train these models on the newly released SKIPP'D dataset and compare our results to its associated baseline model, SUNSET. We consistently outperform SUNSET on sunny images, while underperforming on cloudy images. Adding power logs as an input greatly improves the performance of ViT4SF on cloudy images and has minimal effects on sunny images, proving the importance of such input. These models can be used as a baseline for future studies exploring the use of vision transformers for the task of solar forecasting.

References

Alammar, J. (2018). The Illustrated Transformer. http://jalammar.github.io/illustrated-transformer/.

Augustine, J. A., DeLuisi, J. J., & Long, C. N. (2000). SURFRAD–A national surface radiation budget network for atmospheric research. *Bulletin of the American Meteorological Society*, *81*(10), 2341-2358.

Bonaccorso, G., Fandango, A., & Shanmugamani, R. (2018, December). *Python: Advanced Guide to Artificial Intelligence*. O'Reilly Online Learning. https://www.oreilly.com/library/view/python-advanced-guide/9781789957211/90246d05-866b-474d-834a-9329abfbc488.xhtml

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In International Conference on Learning Representations.

El Alani, O., Abraim, M., Ghennioui, H., Ghennioui, A., Ikenbi, I., & Dahr, F. E. (2021). Short term solar irradiance forecasting using sky images based on a hybrid CNN–MLP model. *Energy Reports*, *7*, 888-900.

Feng, C., & Zhang, J. (2020). SolarNet: A sky image-based deep convolutional neural network for intra-hour solar forecasting. Solar Energy.

Fernandez-Jimenez LA, Muñoz-Jimenez A, Falces A, Mendoza-Villena M, Garcia- Garrido E, Lara-Santillan PM, et al. Short-term power forecasting system for photovoltaic plants. Renew Energy 2012;44:311–7.

Gao, M., Li, J., Hong, F., & Long, D. (2019). Short-term forecasting of power production in a large-scale photovoltaic plant based on LSTM. *Applied Sciences*, *9*(15), 3192.

Goodfellow I. Bengio Y. & Courville A. (2016). *Deep learning*. MIT Press.

Konstantinou, M., Peratikou, S., & Charalambides, A. (2021). Solar Photovoltaic Forecasting of Power Output Using LSTM Networks. *Atmosphere*, *12(1)*, *124*.

Lee, D., & Kim, K. (2019). Recurrent neural network-based hourly prediction of photovoltaic power output using meteorological information. *Energies*, *12*(2), 215.

Marquez, R., Pedro, H. T., & Coimbra, C. F. (2013). Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs. *Solar Energy*, *92*, 176-188.

MathWorks. (n.d.). What is a Convolutional Neural Network?. https://www.mathworks.com/discovery/convolutional-neural-network-matlab.html

Mishra, M. (2020). *Convolutional Neural Networks, explained*. Medium. https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939

Ng, A. (2017). Long Short Term Memory (LSTM). Coursera. https://www.coursera.org/lecture/nlp-sequence-models/long-short-term-memory-lstm-KXoay

Nie, Y., Li, X., Scott, A., Sun, Y., Venugopal, V., & Brandt, A. (2022). SKIPP'D: a SKy Images and Photovoltaic Power Generation Dataset for Short-term Solar Forecasting.

Nie, Yuhao, Zamzam, Ahmed S., & Brandt, Adam. (2021). Resampling and data augmentation for short-term PV output prediction based on an imbalanced sky images dataset using convolutional neural networks. United States. https://doi.org/10.1016/j.solener.2021.05.095

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, *32*.

Pedro, H. T., Larson, D. P., & Coimbra, C. F. (2019). A comprehensive dataset for the accelerated development and benchmarking of solar forecasting methods. *Journal of Renewable and Sustainable Energy*, *11*(3), 036102.

PyTorch. (n.d.). Tanh. https://pytorch.org/docs/stable/generated/torch.nn.Tanh.html

PyTorch. (n.d.). ReLU. https://pytorch.org/docs/stable/generated/torch.nn.ReLU. html

PyTorch. (n.d.). Sigmoid. https://pytorch.org/ocs/stable/generated/torch.nn.Sigmoid.html

Sobri, S., Koohi-Kamali, S., & Rahim, N. A. (2018). Solar photovoltaic generation forecasting methods: A review. *Energy conversion and management*, *156*, 459-497.

Stoffel, T. & Andreas, A. (2015): NREL Solar Radiation Research Laboratory (SRRL): Baseline Measurement System (BMS); Golden, Colorado (Data). National Renewable Energy Laboratory. 10.7799/1052221

Sun, Y., Venugopal, V., Brandt, A.R., 2019. Short-term solar power forecast with deep learning: 762 Exploring optimal input and output configuration. Sol. Energy 188, 730–741. 763 https://doi.org/10.1016/j.solener. 2019. 06.041

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2020, October). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations* (pp. 38-45).

Zhang, A., Lipton, Z., Li, M., & Smola, A. (2021). Dive into Deep Learning. *arXiv preprint arXiv:2106.11342*.

Zhao, X., Wei, H., Wang, H., Zhu, T., & Zhang, K. (2019). 3D-CNN-based feature extraction of ground-based cloud images for direct normal irradiance prediction. *Solar Energy*, *181*, 510-518.

Modeling the Effects of Macroeconomic Policies on Business and Consumer Confidence During the COVID-19 Pandemic

Lucy Lu

Author Background: *Lucy Lu grew up in China and currently attends Shanghai High School International Division in Shanghai, China. Her Pioneer research concentration was in the field of economics and titled "Macroeconomics/Government Policies in Response to the Pandemic."*

Abstract

This paper explores the relationship between the dependent variables of business and consumer confidence and the independent variables of changes in fiscal policy, interest rates, and the number of COVID cases as a percentage of countries' total population during the COVID-19 pandemic in OECD nations. The paper primarily uses correlation-regression analyses to investigate the relationship between the aforementioned variables, and utilizes both cross-sectional and longitudinal data for econometric modeling. The results of this investigation show that while there is a positive correlation between expansionary fiscal policies and business confidence, results yielded for the correlation between business confidence and expansionary monetary policies as well as that between consumer confidence and expansionary fiscal and monetary policies are contrary to expectations. Overall, the paper supports the importance of the consumer confidence index (CCI) and the business confidence index (BCI) as economic indicators.

1. Introduction

Many conventional literature and theories support the stabilizing effects that counter-cyclical monetary and fiscal policies have on the macro-economy (C.W., 2013). In particular, British economist John Maynard Keynes contended that governments should take measures to combat economic downturns through expansionary fiscal policies, whilst American economist Milton Friedman critiqued the Federal Reserve for not enacting monetary policies during the Great Depression (Macroeconomics, n.d.; "Who Was Milton Friedman," 2022). The modern study of macroeconomics now integrates both Keynesian and Monetarist

schools of thought, promoting the use of both expansionary fiscal and monetary policies during periods of recession or depression. The two major discretionary fiscal policy tools include government spending and the tax system, which directly influence government expenditures (G) and consumption (C), respectively. Monetary policies, which are autonomously decided upon by nations' central banks, aim to regulate inflation and unemployment rates by targeting the interest rate and money supply. The major monetary policy tools include open market operations, reserve requirements, and discount rates—all of which can be used to either increase or decrease liquidity. During the COVID-19 pandemic, these policies were used by most—if not virtually all—countries that were affected by the outbreak across the globe.

After the first case of COVID-19 broke out near the end of 2019 in China, many spectators and international institutions feared that COVID-19 might "plunge [the] global economy into the Worst Recession since World War II," as put forth by the World Bank, which also forecasted that "the global economy will shrink by 5.2% [in 2020]" (World Bank Group, 2022). The United Nations also remained pessimistic during the year 2020, accentuating the impact of prolonged lockdowns and protection measures on the service sectors. The United States of America—the country with the world's highest GDP up-to-date—also anticipated that the pandemic would wreak havoc on its economy, because on top of the disruptions to the global supply chain and international trade, "the U.S. economy depends on the optimism of its consumers, which has been shattered by the COVID-19 pandemic" (Ercolano, 2020). While unemployment did surge in most countries during the outbreak, the recession due to COVID-19 only lasted for a mere two months in the USA. Moreover, major European economies were also on their way to recovery by mid-June 2020, suggesting that the duration of the recession induced by COVID-19 was not long.

The quick recovery may be partially attributable to the massive government support packages and central banks' monetary stimuli. The government of the United Kingdom, for instance, launched fiscal stimulus packages and delayed billions in VAT payments to tackle the country's rising unemployment rate (Islam, 2020). Meanwhile, the Bank of England made announcements early on in 2020 about purchasing government debt to increase the money supply. Likewise, the governments and central banks of most other nations also increased spending to prevent further increases in the unemployment rate. Indeed, drastically surging unemployment rates are of top concern in many countries hit hardest by COVID-19, which include but are not limited to Spain and Italy. Hence, due to disruptions to employment and the global supply chain, the global economy might slip into a double-dip recession, with global unemployment expected to rise to more than 200 million ("COVID Crisis to Push Global Unemployment Over 200 Million Mark in 2022," 2021). Still, there are diverging opinions about the future trend of the global economy after the pandemic, as new structural economic models have been devised in recent years following the outbreak of the coronavirus ("The Euro Area's COVID-19 Recession Through the Lens of an Estimated Structural Macro Model," 2021).

Thus, this paper aims to provide insights into economic trends and forecasts following COVID-19 with respect to macroeconomic policies—namely, monetary and discretionary fiscal policies—and the severity of COVID-19, measured in terms of the number of confirmed cases per million, through a cross-

country econometric analysis. Although the aforementioned direct impacts of such policies indicate that they do increase the overall aggregate demand of the economy, household consumption—despite large variations across countries—remains to be the largest component of GDP ("Consumption as Percent of GDP Around the World," n.d.). The major factors that are deterministic of consumption include inflation and cost of living, real incomes, unemployment rate, consumer confidence, and more (Pettinger, 2018). Although these factors are enumerated as separate influences, they are to a great extent interlinked and influence each other simultaneously. Beyond consumer expenditure, business investments are also crucial to economic recovery and deterministic of future economic growth or decline. Therefore, this paper will specifically explore the effects of expansionary macroeconomic policies on consumer and business confidence. In particular, multiple regression analysis will be utilized for the purpose of this paper, with a focus on major European economies within the Organization for Economic Co-operation and Development (OECD).

2. Literature Review

The outbreak of the pandemic was accompanied by drastic increases in national healthcare spending, which may be associated with many problems including but not limited to an accumulation of budget deficits and heightened public debt levels. The high debt levels may increase macroeconomic risks and, as suggested by literature, are overly "expansive and of the wrong form" (Makin & Layton, 2021). Other research analyzed the correlation between changes in GDP and fatality rates, concluding that most countries or regions generally fall under two categories: "large GDP losses and high fatality rates," or "low GDP losses and low fatality rates." Additionally, the research suggests that COVID-19 policies have spillover effects, as the conditions of one country may often have substantial impacts on the "health and economic outcomes" of its neighboring countries (Fernández-Villaverde & Jones, 2020). Consequently, despite the strain that increased healthcare spending places on the government budget, it is also crucial to nations' overall recovery and offers "an opportunity for a 'reset' in countries with weak health financing systems to progress towards universal health coverage" (World Health Organization, 2020).

Similarly, past research also analyzed the effects of expansionary monetary policies either in a previous time period or during the COVID-19 pandemic. A research paper that selected a sample of 37 countries evaluated the effectiveness of monetary policies in stimulating the financial market during the pandemic and found that both conventional and unconventional forms of monetary policy tools have had little influence on the four financial market indicators—including "10-year government bond yields, stock index returns, changes in exchange rates, and growth rates in CDS spreads" (Wei & Han, 2021). On the other hand, another paper examined expansionary monetary policies by looking at changes in interest rates in the USA, and through utilizing an empirical economic model, found that drastic cuts in the federal funds rates have the potential to substantially mitigate the increase in unemployment during a period of economic downturn—specifically, during the COVID-19 pandemic (Cúrdia, 2020). Moreover, it has also been contended that monetary policies are crucial for countries during the pandemic as they are complements of containment

policies. In other words, containment measures should be accompanied by expansionary monetary policies employed by the central bank to "help stabilize economic activity" (Brzoza-Brzezina, Kolasa, & Makarski, 2020). It was also put forth by a study that looked at 11 euro-area countries where central bank asset purchase influences financial markets through a channel known as the "announcement effect" (Moessner & de Haan, 2022). In particular, announcements made by the European Central Bank regarding its asset purchase program—or the Pandemic Emergency Purchase Program (PEPP)—had considerable influence on government bond yields. The influence that monetary policies in the countries investigated by the aforementioned studies can have on unemployment, the financial market, and other economic indicators are also to a large extent intertwined with the indicators that assess consumer and business behaviors, or the CCI and BCI.

Many past studies have statistically analyzed the CCI as an indicator of economic well-being and a potential macroeconomic predictor. For instance, a study that examined the statistical correlation between the CCI and GDP across time in the United States by employing a vector autoregression model (VAR) augments the conclusion of previous literature, which indicate that the CCI is a suitable economic forecaster (Mazurek & Mielcová, 2017). Islam and Mumtaz (2010) also established a long-run positive correlation between consumer sentiment and economic growth in five selected European countries—including the United Kingdom, Germany, France, Denmark, and the Netherlands. The importance of the CCI as an economic indicator was further elaborated upon in Lahiri et al. (2015)'s paper, which points out that consumer spending accounts for approximately "two-thirds of domestic final spending in the USA" and that consumer confidence was found to have a "pervasive effect on all components of aggregate consumption: durables, non-durables, and services" (Lahiri, Monokroussos, & Zhao, 2016). Indeed, the CCI has often been proven and contended to be a crucial economic indicator, and has been within the interest of macroeconomic research.

Certainly, there are also studies that contend otherwise; through econometric modeling, Batchelor and Dua (2003) made the case that the CCI would have been useful for forecasting the 1991 recession because of its special nature, yet would not necessarily have been helpful for predicting economic activity in other years. A paper analyzing the CCI in EU nations also noted that the consumer sentiment indices examined provide "limited information about the future path of household spending" and would only "appear to foreshadow the movement of other major macroeconomic variables" (Cotsomitis & Kwan, 2006). There are also papers that proclaim that confidence factors are crucial for business cycle analysts to make economic forecasts, yet due to the "subjective nature of confidence," the influence that it has on economic behaviors may be limited (Santero & Westerlund, 1996). Indeed, other than economic factors, confidence—particularly consumer confidence—may also be immensely affected by political factors, as argued by Boef and Kellstedt: "politics is important for understanding consumer sentiment beyond what we know from economic conditions" (Boef & Kellstedt, 2004).

Still, even with the ongoing debate revolving around the significance of the CCI as an economic forecaster, most studies—many of which utilize multivariate time series analysis—do suggest that the CCI has some ability to

predict economic trends. Although the aforementioned literature does suggest that the CCI may be under the influence of the political backdrop, it also suggests that people's economic expectations are inextricably intertwined with macroeconomic policies and the future state of the economy.

Further, as it often requires time for consumers of an economy to realize macroeconomic changes, many economists consider the CCI as a lagging economic indicator; however, because consumer confidence may change consumption patterns, other economists consider it to be a leading indicator ("Understanding the Consumer Confidence Index," 2021).

Similarly, the BCI has also been investigated by many economists, and its significance as a potential economic forecaster makes it an interesting topic of study. The BCI is well-known for being a "leading indicator of future output," as it has a "predictive ability for investment growth" and can forecast "investment downturns over 1–3-quarter forecast horizons and for the sign of investment growth over a 2-quarter forecast horizon" (Khan & Updahayaya, 2019). Indicators similar to the BCI—such as "the US Institute for Supply Management's (ISM) manufacturing and Nonmanufacturing Business Activity Index" — are also capable of influencing both stock market regimes and actual industrial production (Çevik, Korkmaz, & Atukeren, 2011). In a study examining consumer and business sentiment in Australia, statistical analyses show that interest rates including "expected and surprise increases in the official cash rate target and related interests" negatively impact consumer confidence, while business confidence has been shown to be less affected by the cash rate target (Kirchner, 2020). The paper also puts forth that there is "only limited evidence for monetary polic[ies] having a perverse signaling effect on sentiment" (Kirchner, 2020). The different findings offered by existing literature in the discipline demonstrate the inherent differences between these indicators in different countries, indirectly pointing to the psychological and cultural factors that also play a role in determining changes in the business and consumer confidence indexes.

More recent papers analyze the CCI and BCI in the context of the COVID-19 pandemic. One study conducted by Teresiene et al. in 2021 specifically analyzed "the impact of the COVID-19 pandemic on consumer and business confidence indicators" in the Eurozone, the United States, and China (Teresiene et al., 2021). Consumer confidence was assessed using the CCI, while business sentiment was assessed using the manufacturing purchasing manager's index (PMI) and services PMI. The study focused on the direct impacts of COVID-19 on the aforementioned indices—with the independent variable being the number of confirmed cases, the number of mortalities due to COVID cases, and the mortality rate of COVID-19 infections—and concluded that the spread of the coronavirus negatively affected the CCI of the USA and China and have mixed effects on the PMIs. Another recent research article investigating the effects of COVID-19 on business confidence in the Eurozone discovered that a decrease in business confidence by one standard deviation also leads to a 5-9% fall in industrial and wholesale and retail trade sectors (Ambrocio, 2022). In all, it has been shown that although there are controversies regarding the CCI and BCI, they are generally significant indicators for the economy and can provide meaningful insights during special time periods such as the recent pandemic.

3. Methodologies

3.1 General Overview of Methodologies

As previously mentioned, both consumer and business confidence levels are crucial determinants of economic well-being. However, due to the lack of an internationally standardized measurement for consumer confidence—or consumer sentiment, which will be used interchangeably with the term consumer confidence for the purpose of this paper—the computation of the Consumer Confidence Index is often conducted with different metrics on a national level. For example, there are two indices for consumer confidence—the University of Michigan's Consumer Sentiment Index and the Conference Board Consumer Confidence Index—in the USA alone. For many European nations such as France, Germany, Italy, and the UK, consumer sentiment is measured through approximately 2,000 telephone surveys conducted by INSEE, GFK, ISAE, and GFK, respectively (Golinelli and Parigi, 2003). As these data are obtained by different economic institutions, an accurate conversion would be difficult to achieve. However, the OECD offers a database that adjusts the CCI for each of its member nations; moreover, OECD economies generally have many commonalities with one another and are most economically developed. Hence, this essay will primarily analyze confidence levels in OECD member nations as the converted CCI and BCI indices are readily available, and the omitted variable bias may not be as prevalent, as psychological and cultural factors may play a less significant role in determining consumer and business confidence levels.

For the purpose of this paper, fiscal policies will be discussed in terms of additional government spending and foregone tax revenue. Although the tax and spending multiplier are generally computed with different formulas, the two measures will be regarded simultaneously as one single independent variable—as the precise multiplier for each country cannot be attained and would only further complicate the model. As previously mentioned, there are many monetary tools that the central banks of nations utilize. However, because the effects of the usage of monetary policies are generally assessed by changes in interest rates and money supply—or primarily M2—data for the two variables were initially considered (Benge, 2021). Yet since there is also an inverse relationship between interest rates and money supply, an issue of collinearity arises. Ultimately, the change in interest rates is selected as the independent variable to assess monetary policy usage, mainly because interest rates affect business investments more directly. Moreover, as most OECD member nations are European countries, using the short-term interest rate would be implausible as most countries would then have an equal interest rate as set by the European Central Bank (ECB). The sources for all data obtained for the following econometric analysis are stats.oecd.org and data.oecd.org, with the exception of that for fiscal policy implementation, for which the data source is the International Monetary Fund (IMF).

3.2 Methodologies for Assessing BCI in Relation to Changes in Fiscal Measures, Long-Term Interest Rates, and the Number of COVID Cases as a Percentage of Population

The technique used to investigate the correlation between the BCI and the three independent variables—fiscal policies, long-term interest rates, and the number of COVID cases as a percentage of the national population—will be a multiple regression analysis of the data collected. While the analysis aims to take all 38 OECD countries into consideration, five countries—namely, Costa Rica, Estonia, Iceland, Norway, and Turkey—are omitted from the sample due to the lack of either one or more than one piece of data that precludes it from being useable for the regression. Furthermore, as the dependent variables are measured in terms of percentage changes, two time periods must be selected for the calculation to be done; since COVID broke out in China near the end of 2019, it likely had not had a significant effect on the consumer and business confidence of most European or Western nations by 2019. Most effective macroeconomic measures to combat the receding economy during the coronavirus outbreak also primarily took place before October 2021. Therefore, the time nodes used for this paper are January 2020 and October 2021—that is, the percentage change for all variables is calculated as follows:

$$\%\Delta x = \frac{x_{Oct\ 2021} - x_{Jan\ 2020}}{x_{Jan\ 2020}}$$

or

$$\%\Delta y = \frac{y_{Oct\ 2021} - y_{Jan\ 2020}}{y_{Jan\ 2020}}.$$

The selection of October 2021 can also be justified by the fact that many central banks have been gradually increasing interest rates and withdrawing from expansionary monetary policies; for example, the Fed "raised the rate to a range of 0.25% to 0.50%" in March of 2022 and "hike[d] interest rates by 0.75 percentage point[s] for [the] second consecutive time" during July in an effort to combat inflation. To prevent inaccuracies, October 2021 is chosen as the second time node and considered an approximate end to the global economic crisis induced by the coronavirus outbreak.

After conducting a multiple regression analysis, the results will be summarized in the format of an equation:

$$y_t = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

The α denotes the intercept of the graph, and the $\beta_1, \beta_2, and\ \beta_3$ denote the different correlation coefficients for $x_1, x_2, and\ x_3$, respectively. The multiple regression model designed to analyze the dependent variable of the business confidence index, $x_1, x_2, and\ x_3$ represent net changes in government spending and foregone revenue as a percentage of GDP, changes in interest rates, and the number of COVID cases as a percentage of the population, respectively. u is the

term of the multiple regression equation that embodies the random disturbance that is not accounted for by the independent variables.

3.3 CCI Analysis Methodologies

The consumer confidence index is analyzed with respect to the same independent variables used in the multiple regression analysis for BCI. Indeed, the results of multiple regression analysis will still be in place and attached to the paper. However, due to the accompanying time lag with the consumer confidence index, other means to assess the effects of COVID and macroeconomic policies used during the pandemic must be introduced. As the number of COVID cases as a percentage of the population is not a macroeconomic policy, there will not be an additional section specifically devoted to it. Instead, the new methods of analysis will only be used for the independent variables of fiscal measures and changes in interest rates.

The correlation between changes in interest rates and the CCI will be modeled not through a multiple regression analysis, but rather through a time series analysis; specifically, a simplified auto-regressive model will be used for the construction of the model. Whilst time series analysis with multiple variables—with time being considered as a variable and therefore requiring an added dimension—requires complex panel data analysis, the simpler version of the model will conduct separate time series analyses for each of the OECD member countries. In other words, the same analysis will be conducted multiple times so that the cultural and psychological factors will not become confounding variables to the results. Although ideally 38 auto-regressive models can be constructed for all 38 member nations, the partially lacking data only allows for the test to be conducted on 25 of the member nations. The list of countries and raw data are attached in the appendix section of the paper.

4. Analysis of General Statistics Pertinent to the CCI & BCI During the COVID-19 Pandemic

4.1 Analysis of General Trends for Changes in the CCI and BCI

The OECD database shows a general trend similar to that of quarterly GDP and a smaller standard deviation as compared to the CCI. To analyze the data regarding consumer and business confidence indices holistically, this section will utilize graphs generated from data.oecd.org

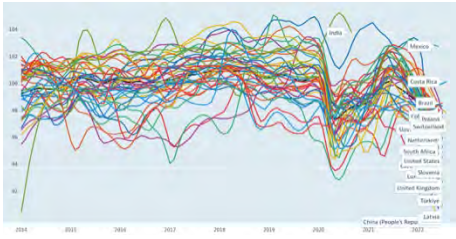


Figure 1. *Consumer Confidence Index of OECD Nations Over Time (Source: data.oecd.org)*

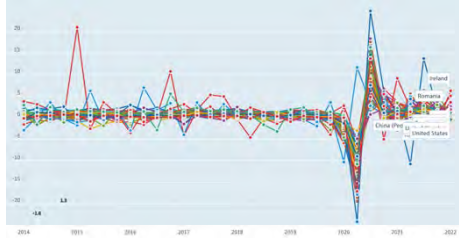


Figure 2. *Quarterly GDP by Country (Source: data.oecd.org)*

that show general increases or dips for the aforementioned confidence levels.

Figure 1 shows the CCI of all OECD nations from 2014 to 2022 and indicates that while there is a general trend for consumer confidence levels, there is also a high standard deviation from the mean. The extent of fluctuation at distinct time periods also varies greatly by country; for instance, the concurrent dip in consumer confidence at the beginning of 2020 when the coronavirus first became widespread differs in magnitude for different OECD nations. Reasonably, the severity of the outbreak in these nations as commonly measured by the number of cases and deaths accounts at least partly for the disparity in the decrease in CCI. While other factors such as culture and psychological factors may also constitute a considerable portion of the difference observed, macroeconomic policies may also be exerting a significant influence on the magnitude of the oscillations. Furthermore, the general trend for CCI, as shown in Figure 1, also corresponds with changes in quarterly GDP, as indicated in Figure 2. These similarities insinuate that macroeconomic policies not only greatly impacted the major economic indicator of GDP but also consumer and business confidence.

As the Business Confidence Index (BCI)—which is also converted to the

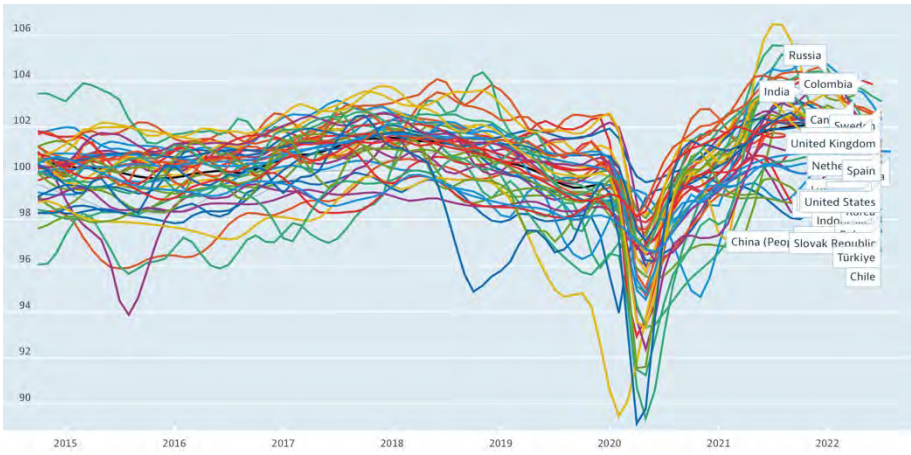


Figure 3. *Business Confidence Index of OECD Nations Over Time (Source: data.oecd.org)*

same base by the OECD—greatly influences employment levels, assessing the potential relationship between macroeconomic policies and the BCI during a special timeframe such as COVID-19 may have meaningful economic implications.

4.2 Summary of Data Using Descriptive Statistics

Table 1. *Average Changes in CCI, BCI, Interest Rates, Gov't Spending, and # of COVID cases*

	CCI (Jan 2020)	CCI (Oct 2021)	BCI (Jan 2020)	BCI (Oct 2021)	Interest Rate (Jan 2020)	Interest Rate (Oct 2021)
Mean	100.67	100.47	99.91	102.12	1.01	1.43
Median	100.71	100.60	100.03	102.41	0.37	0.39
Maximum	103.17	103.18	102.59	104.73	6.85	8.40
Minimum	96.00	97.69	96.44	98.25	-0.70	-0.20
Range	7.17	5.49	6.15	6.48	7.55	8.60
Standard Deviation	1.63	1.28	1.45	1.73	2.21	2.10

Table 2. *Average CCI, BCI, and Interest Rates (Based on Author's Own Calculations from Data Provided by data.oecd.org)*

	Change in CCI	Change in BCI	Change in Gov't Spending	Change in Interest	% case of population
Mean	-0.0018	0.0222	10.6	0.413	0.2613
Median	0.0667	0.0864	9.6	0.190	0.0889
Maximum	0.0297	0.0797	25.5	3.034	6.0207
Minimum	-0.0370	-0.0066	0.7	-0.380	0.0003
Range	0.0667	0.0864	24.8	3.414	6.0204
Standard Deviation	0.0172	0.0227	5.6	0.708	1.0350

Table 3. *Bounds for Finding Outliers*

	BCI	CCI	Change in gov't spending	Change in Interest	% case of population
Q1	0.0077	-0.0109	6.4	0.053	0.0105
Q3	0.0354	0.0079	15.2	0.570	0.1189
IQR	0.0277	0.0188	8.8	0.517	0.0785
lower bound	-0.0338	-0.0390	-6.8	-0.723	-0.0772
upper bound	0.0769	0.0360	28.4	1.346	0.2366

The tables above provide a summary of the data used for the multiple regression analysis. Outliers are determined based on the lower and upper bounds. The formulas used to calculate the lower and upper bounds are as follows:

$$Lower\ bound = Q1 - 1.5 * IQR$$
$$Upper\ bound = Q3 + 1.5 * IQR$$

Despite the relatively high standard deviation that the dataset contains, there are no outliers to any of the dependent and independent variables used for the multiple regression analysis. As the mean change in the BCI is much higher than that in the CCI, cross-sectional analysis of the effects of the independent variables on BCI will likely be more meaningful. This also provides more justification as to why—based on the standard deviation calculated in Figure 4—a longitudinal analysis that analyzes different countries separately may prove to be more insightful.

4.3 Constructing a Linear Regression Model for the Correlation between the CCI and BCI

Moreover, because consumer confidence usually drives economic activity, it would also be reasonable to infer that consumer sentiment influences business confidence to a great extent. For instance, a time series analysis analyzing the Australian economy over the span of the past few decades suggests that there is indeed a correlation between the two variables and that consumer confidence is potentially driving business confidence, with the divergence between the two trendlines primarily arising from "politics, debt, technical [discrepancies] and capacity utilization." Therefore, it also becomes crucial to first examine the relationship between the consumer and business confidence indices before diving into the extent to which they are influenced by COVID severity and macroeconomic policies. For this purpose, a linear regression test is used to investigate the correlation between the two variables statistically. The data used for the below linear regression analysis is also obtained from the aforementioned sources (data.oecd.org), and the percentage change in both
indicators are as described in the methodologies section.

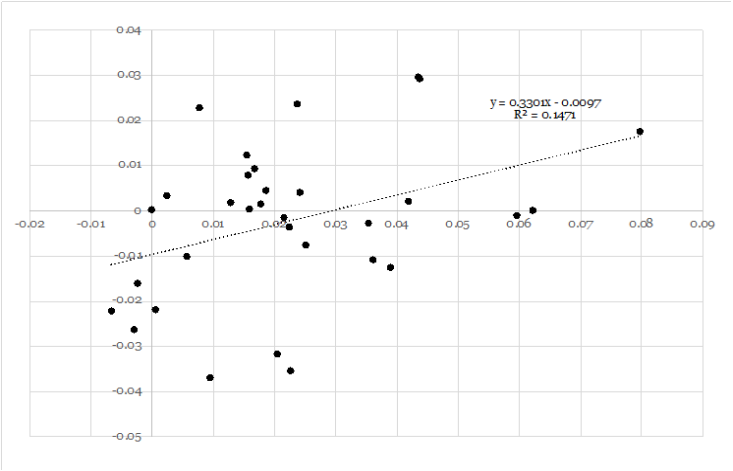


Figure 4. *Linear Regression Results for the Correlation Between the CCI and the BCI (Based on the author's own calculations from data provided by data.oecd.org)*

Each point in Figure 4 represents a country's combination of changes in consumer and business confidence levels, with the x-coordinates denoting the change in consumer sentiment and the y-coordinates denoting the change in business confidence. As shown on the scatter plot, there exists a correlation between changes in consumer and business confidence levels in the 33 OECD countries investigated. When applying statistical tests, the results also prove to be statistically significant. Below is a summary of the results of the linear regression analysis:

Table 4. *Summary of Linear Regression Results (Based on the author's own calculations from data provided by data.oecd.org)*

Regression Statistics	
Multiple R	0.38350533
R Square	0.14707634
Adjusted R Square	0.11864555
Standard Error	0.01894256
Observations	32

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.001856227	0.00185623	5.17313605	0.03025505
Residual	30	0.010764616	0.00035882		
Total	31	0.012620843			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.02370837	0.003375243	7.02419807	8.3077E-08
Change in CCI (x)	0.4455419	0.195889699	2.27445291	0.03025505

	Lower 95%	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.01681521	0.030601538	0.01681521	0.03060154
Change in CCI (x)	0.04548176	0.845602031	0.04548176	0.84560203

The above statistical analysis and summary output show that the correlation has a p-value of approximately 0.0303, which is below 0.05 and therefore statistically significant. The value of R-squared is around 0.147, which suggests that 14.7% of changes in the BCI can be explained by changes in CCI. The results can also be summarized in the form of a simple equation:

$$BCI = \alpha + \beta \cdot CCI + u$$

In the equation, the value of α is approximately 0.0237, whilst the value of β is around 0.446. From an economics standpoint, the results make sense: the consumer confidence index reflects consumers' outlooks on and expectations for the present and future state of the economy. Due to the partially lagging and partially leading nature of the consumer confidence index, consumer sentiment also provides insights into consumer spending, which may in turn impact business confidence. In other words, increases in consumer confidence may stimulate increases in business confidence, as shown through the positive coefficient of an estimated 0.446 as shown in Figure 7. However, since the correlation modeled above also have a standard error—or the average distance of the observed value from the predicted value modeled by the regression line—of 0.196, the two variables will still be analyzed separately. As the two variables also influence two different components of GDP, conducting multiple regression analyses on both variables will also yield more meaningful conclusions regarding the effectiveness of macroeconomic policies during special time periods such as the COVID-19 pandemic.

5. Exploration

5.1 BCI Multiple Regression Analysis

As previously discussed in the methodologies section, a multiple regression analysis using Microsoft Excel is conducted. Unlike the CCI—which will be analyzed in a later section—the BCI is oftentimes not considered a lagging indicator, but rather a leading indicator that can be used to forecast future economic development.

Table 5. *Summary of Multiple Regression Results for BCI (Based on the author's own calculations from data provided by data.oecd.org)*

Regression Statistics	
Multiple R	0.49447343
R Square	0.24450398
Adjusted R Square	0.16634922
Standard Error	0.01832654
Observations	33

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	0.003152193	0.00105073	3.12845915	0.040822885
Residual	29	0.009740003	0.00033586		
Total	32	0.012892196			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.00342304	0.007762756	0.44095674	0.6625143
Fiscal Policy Changes in Long-Term Interest Rate	0.00128437	0.000596195	2.15427315	0.03966179
% Case of population	0.01144403	0.004641325	2.46568239	0.01983265
	0.00205345	0.003218064	0.63810129	0.52841479

	Lower 95%	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-0.0124536	0.019299658	-0.0124536	0.01929966
Fiscal Measures Changes in Long-Term Interest Rate	6.5011E-05	0.002503723	6.5011E-05	0.00250372
% Case of population	0.00195146	0.02093661	0.00195146	0.02093661
	-0.0045282	0.008635132	-0.0045282	0.00863513

The econometrics model for the multiple regression analysis is as follows:

$$BCI_t = \alpha + \beta_1 \cdot Fiscal\ Measures + \beta_2 \cdot Changes\ in\ Long - Term\ Interest\ Rates + \beta_3 \cdot Number\ of\ Cases\ As\ A\ \%\ of\ Population + u$$

Where:

$$\beta_1 = 0.00128437$$
$$\beta_2 = 0.1144403$$
$$\beta_3 = 0.00205345$$

As indicated by the above results, there is a statistically significant positive correlation between the expansionary fiscal policies used and increases in business confidence over the 22-month period from January 2020 and October 2021. The statistical significance of the correlation also makes logical sense and adheres to the economic theory; while additional government spending increases government expenditures, reduced taxes simultaneously boost the investment component of countries' GDP. Therefore, with a positive correlation coefficient of approximately 0.00128, indicating that a one percent change in fiscal policies as measured in additional government spending and foregone revenue corresponds with a 0.00128% increase in BCI from a cross-sectional point of view. Although the corresponding change in BCI is incremental, it is important to note that many countries have applied expansionary fiscal policies that amount to more than 20% of their GDP (IMF data pdf).

Moreover, there seems to be a stronger correlation between the independent variable and dependent variable of changes in interest rates, which assesses the magnitude and effectiveness of expansionary monetary policy usage. While the results are statistically significant with a p-value of 0.0198, the correlation coefficient is, contrary to expectations, positive. According to conventional theory, reduced interest rates stimulate business investment and confidence because they diminish the cost of borrowing and incentivize the launching of new projects. Interestingly, this is not the case for the regression analysis above. Still, the results obtained through the multiple regression could be explained by the leading nature of the BCI as an economic indicator. As the BCI foreshadows the upcoming economic trend, a low value of the BCI potentially implicates an economic downturn. Therefore, decreases in the BCI may have served as cues for the central bank to lower interest rates in order to boost business activity. Fundamentally, a multiple regression analysis is limited to showing correlation and not causation. Hence, the positions of the independent and dependent variables could be switched, and the statistical significance and results would remain the same. Perceiving business sentiment as a leading indicator that to some extent provided guidance to central banks, therefore, offers a plausible explanation for the unanticipated positive correlation. In essence, a one percent change in interest rate corresponds to a 0.0114% change in business confidence because central banks of the OECD member nations considered may have responded to the signals provided by the BCI or other leading indicators that displayed similar trends.

There also does not seem to be a statistically significant correlation between the number of cases expressed as a percentage of the country's total population and the net change in the BCI in the sample of the 34 OECD countries selected. While in theory, business confidence should be negatively impacted by the severity of the outbreak, the lack of a distinct relationship could also be explained through the relatively small sample size and external factors that cannot be accounted for by econometric analysis—such as cultural differences. Still, the results show that the magnitude of COVID might not play a large role in determining business confidence. The results can be compared and contrasted with a study aforementioned in the literature review section of the paper. Referring back to Teresiene et al. (2021), the longitudinal correlation-regression analysis using different timeframes for variable obtention demonstrated that in

the Eurozone, business sentiment seems to be statistically significantly impacted by the fatality rate, "the Eurozone PMI in the manufacturing sector demonstrated a positive reaction to the increase in COVID-19 cases and deaths both in-country and globally." As the OECD encompasses many Eurozone nations, the United States, and a few other nations across the globe, analysis regarding the effects of the spread of COVID-19 on business sentiment in the Eurozone may share many similarities with that of the selected OECD member nations. The results to some extent corroborate Teresiene et al. (2021), as the spread of COVID-19 in the above analysis is calibrated by the percentage of the population affected by COVID, which also does not seem to have had a significant negative impact as discussed in the conclusion of Teresiene et al. (2021).

5.2 CCI Multiple Regression Analysis and Auto-Regressive Analysis

As the CCI is often considered a lagging indicator, the time frame with which the data are obtained and considered should be different from that for the BCI. Multiple regression analysis is still conducted for the CCI, although the change in CCI is calculated by subtracting the CCI recorded for July 2020 from that recorded for October 2021. The percent change would then be calculated by dividing that difference by the CCI recorded for July 2020. In simpler terms, the change in CCI can be calculated as follows:

$$\%\Delta CCI = \frac{CCI_{Oct\ 2021} - CCI_{July\ 2020}}{y_{July\ 2020}}$$

Still, the data used for the dependent variables will be exactly the same as that used for the BCI model. The results of the cross-sectional data are shown as follows:

Table 6. *Summary of Multiple Regression Results for CCI (Based on the author's own calculations from data provided by data.oecd.org)*

Regression Statistics	
Multiple R	0.44524306
R Square	0.19824139
Adjusted R Square	0.11530084
Standard Error	0.01946797
Observations	33

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	0.00271763	0.00090588	2.39016254	0.089097678
Residual	29	0.01099105	0.000379		
Total	32	0.01370868			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.03326796	0.00824624	4.03431861	0.000364438
Fiscal Policy	-0.001147	0.00063333	-1.811133236	0.080490984
Change in Interest Rates	0.00051664	0.00341849	0.151131635	0.880918286
% Case of Population	0.00816126	0.0049304	1.6552949	0.108648534

	Lower 95%	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.01640251	0.05013342	0.016402505	0.050133415
Fiscal Policy	-0.0024423	0.00014826	-0.002442341	0.00014826
Change in Interest Rates	-0.006475	0.00750825	-0.006474962	0.007508247
% Case of Population	-0.0019225	0.01824506	-0.001922534	0.018245062

The econometrics model for the above multiple regression for CCI is hence:

$$CCI_t = \alpha + \beta_1 \cdot Fiscal\ Measures + \beta_2 \cdot Changes\ in\ Long - Term\ Interest\ Rates + \beta_3 \cdot Number\ of\ Cases\ As\ A\ \%\ of\ Population + u$$

Where:

$$\beta_1 = -0.001147$$
$$\beta_2 = 0.00051664$$
$$\beta_3 = 0.00816126$$

As discussed in the table above, both changes in interest rates and differences in the number of cases as a percent of the population are not statistically significant. There are two possible explanations for this, and the first is that monetary policies as well as the severity of COVID simply do not impact or have a very minimal, if not negligible, impact on consumer sentiment. The second plausible reason is that the part played by cultural and psychological factors and differences is too immense that the results cannot be significant when analyzed cross-sectionally.

While theoretically, expansionary fiscal policy will instigate increases in consumer confidence as increases in government spending and reductions in taxes are aimed at stimulating economic recovery or growth, such is not the case above. Rather, the coefficient for the correlation between CCI and changes in government spending and foregone revenue is negative. This may be a result of the vast number of external variables, especially given the relatively small sample size that the regression analysis uses. Still, this phenomenon would be difficult to explain if the model is not reversed: hence, a revised model that inverts the x and y variables is devised. Since the above multiple regression shows that the number of cases as a percentage of the country's total population as a variable does not affect consumer sentiment, it can be omitted from the revised version of the

regression model. Hence, a linear regression can be conducted as follows:

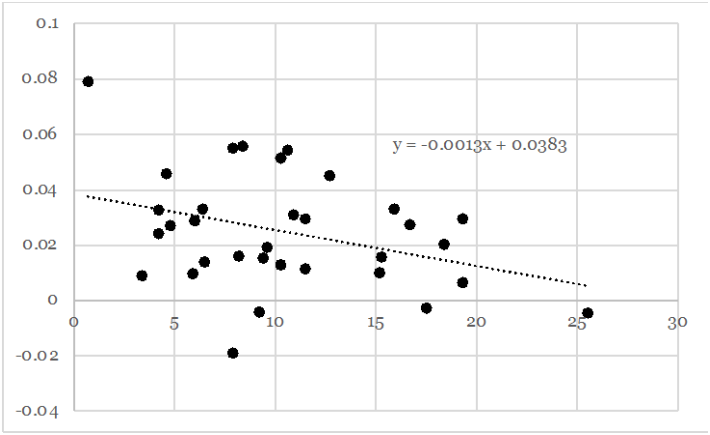


Figure 5. *Linear Regression Results (Model) for CCI and Fiscal Measures (Based on the author's own calculations from data provided by data.oecd.org)*

As indicated in the linear regression above, there is a distinct downward trend, signifying a negative linear correlation between the two aforementioned variables. The results can be interpreted as the CCI being indicative of the economy at a given time period, and a change in government spending can be primarily understood as a means to ameliorate the current economic circumstances. As there is a high opportunity cost of imposing expansionary fiscal policies—including increased government budget deficit and debt, decreased tax revenue for public projects, and potential crowding-out effects——countries that are not experiencing drastic economic downturns will often not opt for radical expansionary fiscal measures. Therefore, the above regression analysis provides indirect evidence that shows CCI is a valuable indicator of a nation's economic state and can be used to forecast certain economic phenomena. Since fiscal policies are not currently based on consumer and business sentiment, the coincidental correlation shows that CCI may indeed be an accurate indicator that can potentially be taken into account by policymakers or economists. The results of the above linear regression diagram also turn out to be statistically significant.

Table 7. *Linear Regression Results (Model) for CCI and Fiscal Measures After Reversing X- and Y- Variables (Based on the author's own calculations from data provided by data.oecd.org)*

Regression Statistics	
Multiple R	0.34989504
R Square	0.12242654
Adjusted R Square	0.09411772
Standard Error	5.33469088
Observations	33

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	123.07569	123.075695	4.324678031	0.045923985
Residual	31	882.22673	28.4589267		
Total	32	1005.3024			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	12.889157	1.45807065	8.83987137	5.58957E-10
CCI	-94.751996	45.5629123	-2.079586	0.045923985

	Lower 95%	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	9.91540228	15.8629117	9.91540228	15.86291167
CCI	-187.67817	-1.8258233	-187.67817	-1.82582333

As suggested by the ANOVA analysis above, the P-value of the linear regression conducted is less than 0.05, and the r-squared value is approximately 0.122. This suggests that the correlation is most likely not coincidental and that changes in CCI potentially account for 12.2% of changes in fiscal policies.

Still, another method should also be used to analyze the effects of monetary policy on consumer sentiment. As opposed to cross-sectional analysis, the following analysis will be longitudinal. Since monthly data from January 2020 through July 2022 are collected, there will be 31 data points for each of the 25 countries analyzed. The details regarding the model have been laid out in the methodologies section of the paper, whilst the raw data are attached in the appendix section.

Through the construction of a simplified auto-regressive model using Microsoft Excel, it is found that, longitudinally, changes in long-term interest rates do have a strong, statistically significant positive correlation with the CCI in some of the countries examined. In particular, it was found that out of 25 countries examined, seven countries—including the USA, the UK, Canada, Denmark, Poland, and Sweden—demonstrated this statistical significance. Interestingly, it is found that the statistically significant results are all found in the regression models' first orders, meaning that the CCI changes accordingly with the interest rate.

Table 8. *Summary of Results for the Auto-Regressive Model (Based on the author's own calculations from data provided by data.oecd.org)*

Country	**Coefficient (first-order: no time lag)**	**P-value**
USA	1.932373111	0.00469629
UK	6.66205355	0.01356378
Canada	5.22917427	0.00076771

Country	Coefficient (first-order: no time lag)	P-value
Czech Republic	3.25123905	0.09838354
Denmark	1.68777967	0.09115574
Poland	4.09026078	0.00266442
Sweden	4.99782031	0.05846926

Similar to the BCI, there is an unexpectedly strong positive correlation between the monthly CCI and long-term interest rate projections. The explanation for this is hence similar to that given for the positive correlation between the BCI and long-term interest rates: the CCI is indicative of the current or past state of the economy. In other words, the well-being of a nation's economic state can be partially calibrated by the CCI, especially in the USA, Canada, and Poland where the significance level, or P-value, of the correlations is below 0.01. For instance, when the economy starts to recover from the downturned induced by the COVID-19 pandemic, central banks may opt to increase the interest rates back to pre-pandemic levels or even higher due to the recently surging inflation occurring simultaneously in multiple parts of the world.

6. Conclusion

In conclusion, this paper explored the relationship between the dependent variables of business and consumer confidence and the independent variables of changes in fiscal policy, interest rates, and the number of COVID cases as a percentage of countries' total population. It is hypothesized that both the CCI and BCI will adjust accordingly to the expansionary fiscal and monetary policies used during the pandemic and that both will also be negatively influenced by COVID, showing significant decreases if there are a large number of cases.

The results of the multiple regression analyses, however, suggest otherwise. The only statistically significant correlation modeled that conforms to conventional theory and expectations is that between changes in BCI and fiscal measures. However, despite the statistically positive correlation between changes in the BCI and fiscal measures, the changes that fiscal measures induce on the BCI are only incremental.

Other results that are statistically significant have all proven to be contrary to expectations. In particular, there is a strong, positive correlation between changes in interest rates and BCI, which means that expansionary monetary policies correspond to decreases in BCI. Moreover, the paper also finds that there is a negative correlation between changes in the CCI and additional spending or foregone revenue, which suggests that expansionary fiscal policies also corresponded to decreases in the CCI, as revealed in a cross-sectional analysis. It is also found through a longitudinal analysis that investigates the correlation between the changes in CCI and interest rates by country that seven of the 25 selected OECD countries (these countries were selected due to the readily available data on the OECD online database) showed a positive correlation between interest rates and the CCI. Since correlation does not equate to causation, a plausible and highly likely explanation for these unpredicted and to

some extent surprising findings is that both confidence indicators—especially the CCI, as it is a lagging indicator—reflect the state of the economy. Therefore, whilst policy-makers do not necessarily rely upon the CCI and BCI when making macroeconomic decisions, such as conducting some form of fiscal or monetary policy,the coincidental positive correspondences between the BCI and interest rates and the CCI, fiscal policies, and interest rates exemplify the usefulness of the CCI and BCI as indicators of economic well-being. Ultimately, further studies that examine consumer and business sentiment should be conducted, as these two variables may offer meaningful insights into nations' economic states, especially during periods when countries from across the globe are suffering from an economic crisis prompted by an event like a pandemic.

References

Ambrocio, G. (2022). Euro-area business confidence and COVID-19. *Applied Economics*, *54*(43), 4915–4929. https://doi.org/10.1080/00036846.2022.2038777

Batchelor, R., & Dua, P. (2021). Improving macro-economic forecasts: The role of consumer confidence. *International Journal of Forecasting*, *14*, 71–81.

Benge, V. A. (2021, November 20). The Effects of Monetary Policy. Retrieved September 15, 2022, from https://bizfluent.com/13635957/the-effects-of-monetary-policy

Boef, S. D., & Kellstedt, P. M. (2004). The Political (And Economic) Origins of Consumer Confidence. *American Journal of Political Science*, *48*(4), 633–649. Retrieved from https://www.jstor.org/stable/1519924

Brzoza-Brzezina, M., Kolasa, M., & Makarski, K. (2020). Monetary Policy and COVID-19. *International Journal of Central Banking*, 41–80.

Çevik, E., Korkmaz, T., & Atukeren, E. (2011). Business confidence and stock returns in the USA: a time-varying Markov regime-switching model. *Applied Financial Economics*, *22*(4), 299–312. https://doi.org/10.1080/09603107.2011.610742

Consumption as percent of GDP around the world. (n.d.). Retrieved September 15, 2022, from https://www.theglobaleconomy.com/rankings/consumption_gdp/#:%7E:text=Household%20consumption%20is%20about%2060%20percent%20of%20GDP,of%20GDP%20to%20over%2080%20percent%20of%20GDP.

Cotsomitis, J. A., & Kwan, A. C. C. (2006). Can Consumer Confidence Forecast Household Spending? Evidence from the European Commission Business and Consumer Surveys. *Southern Economic Journal*, *72*(3), 597–610. Retrieved from https://www.jstor.org/stable/20111835

COVID crisis to push global unemployment over 200 million mark in 2022. (2021, June 4). Retrieved September 15, 2022, from https://news.un.org/en/story/2021/06/1093182

Cúrdia, V. (2020). Mitigating COVID-19 Effects with Conventional Monetary Policy. *FRBSF Economic Letter*. Retrieved from https://www.researchgate.net/publication/340610040

C.W. (2013, October 17). Learning to (push against the) cycle. Retrieved September 15, 2022, from https://www.economist.com/free-exchange/2013/10/17/learning-to-push-against-the-cycle

Ercolano, P. (2020, March 16). Recession appears "inevitable" amid COVID-19 crisis. Retrieved September 15, 2022, from https://hub.jhu.edu/2020/03/16/coronavirus-recession-q-and-a/

The euro area's COVID-19 recession through the lens of an estimated structural macro model. (2021, September 8). Retrieved September 15, 2022, from https://cepr.org/voxeu/columns/euro-areas-covid-19-recession-through- lens-estimated-structural-macro-model

FERNÁNDEZ-VILLAVERDE, J., & Jones, C. I. (2020). Macroeconomic Outcomes and COVID-19. *Brookings Institution Press*, 111–146. Retrieved from https://www.jstor.org/stable/10.2307/27059299

Golinelli, R., & Parigi, G. (2003). What is this thing called confidence? A comparative analysis of consumer confidence indices in eight major countries. *Banca D'Italia*.

IMF Fiscal Affairs Department. (2022). Database of Fiscal Policy Responses to COVID-19 [Dataset]. In *Fiscal Monitor Database of Country Fiscal Measures in Response to the COVID-19 Pandemic*. International Monetary Fund. https://www.imf.org/en/Topics/imf-and-covid19/Fiscal-Policies-Database-in-Response-to-COVID-19

Islam, F. (2020, March 20). Coronavirus recession not yet a depression. Retreived September 15, 2022, from https://www.bbc.com/news/business-51984470

Islam, T. U., & Mumtaz, M. N. (2016). Consumer Confidence Index and Economic Growth: An Empirical Analysis of EU Countries. *EuroEconomics*, (2(35)).

Khan, H., & Upadhayaya, S. (2019). Does business confidence matter for investment? *Empirical Economics*, *59*, 1633–1665. Retrieved from https://link.springer.com/article/10.1007/s00181-019-01694-5

Kirchner, S. (2020). The Effect of Changes in Monetary Policy on Consumer and Business Confidence. *Australian Economic Review*, *53*(1), 118–125. https://doi.org/10.1111/1467-8462.12365

Lahiri, K., Monokroussos, G., & Zhao, Y. (2016). Forecasting Consumption. *Journal of Applied Econometrics*, *31*(7), 1254–1275. Retrieved from https://www.jstor.org/stable/10.2307/26609675

Macroeconomics. (n.d.). Retrieved September 15, 2022, from https://www. https://www.britannica.com/topic/macroeconomics

Makin, A. J., & Layton, A. (2021). The global fiscal response to COVID-19: Risks and repercussions. *Economic Analysis and Policy*, *69*, 340–349. https://doi.org/10.1016/j.eap.2020.12.016

Mazurek, J., & Mielcová, E. (2017). Is consumer confidence index a suitable predictor of future economic growth? An evidence from the USA. *E+M Ekonomie a Management*, *20*(2), 30–45. https://doi.org/10.15240/tul/001/2017-2-003

Moessner, R., & de Haan, J. (2022). Effects of monetary policy announcements on term premia in the euro area during the COVID-19 pandemic. *Finance Research Letters*, *44*, 102055. https://doi.org/10.1016/j.frl.2021.102055

OECD (2022), "Business confidence index (BCI)" (indicator), https://doi.org/10.1787/3092dc4f-en (accessed on 15 September 2022).

OECD (2022), "Consumer confidence index (CCI)" (indicator), https://doi.org/10.1787/46434d78-en (accessed on 15 September 2022).

Organization for Economic Co-operation and Development. (n.d.). *OECD Statistics* [Dataset]. OECD.Stat. https://stats.oecd.org

Pettinger, T. (2018, January 26). Causes of Consumer Spending. Retrieved September 15, 2022, from https://www.economicshelp.org/blog/396/economics/consumer-spending-its-causes-and-effects/

Santero, T., & Westerlund, N. (1996). Confidence Indicators and Their Relationship to Changes in Economic Activity. Organisation for Economic Co-operation and Development, 170.

Teresiene, D., Keliuotyte-Staniuleniene, G., Liao, Y., Kanapickiene, R., Pu, R., Hu, S., & Yue, X. G. (2021). The Impact of the COVID-19 Pandemic on Consumer and Business Confidence Indicators. *Journal of Risk and Financial Management*, *14*(4), 159. https://doi.org/10. 3390/jrfm 14040159

Understanding the Consumer Confidence Index. (2021, August 28). Retrieved September 15, 2022, from https://www.investopedia.com/insights/understanding-consumer-confidence-index/

Wei, X., & Han, L. (2021). The impact of COVID-19 pandemic on transmission of monetary policy to financial markets. *International Review of Financial Analysis*, *74*, 101705. https://doi.org/10.1016/j.irfa.2021. 101705

Who Was Milton Friedman? (2022, May 4). Retrieved September 15, 2022, from https://www.investopedia.com/terms/m/milton-friedman.asp

World Bank Group. (2022, January 14). COVID-19 to Plunge Global Economy into Worst Recession since World War II. Retrieved September 15, 2022, from https://www.worldbank.org/en/news/press-release/2020/06/08/covid-19-to-plunge-global-economy-into-worst-recession-since-world-war-ii

World Health Organization. (n.d.). WHO Coronavirus (COVID-19) Dashboard [Dataset]. In *Data Information*. https://covid19.who.int/data

World Health Organization. (2020). COVID-19: Near- and medium-term implications for health spending. *World Health Organization*.Retrieved from https://www.jstor.org/stable/resrep30130.9

Offshoring and the Coronavirus Pandemic

Selina Song

Author Background: *Selina Song grew up in the United States and currently attends Irvington High School in Fremont, California in the United States. Her Pioneer research concentration was in the field of economics and titled "International Economics."*

Abstract

The first coronavirus outbreak was reported in December 2019 and by March 2020, the disease had spread throughout the world and become a pandemic of unprecedented scale that would change people's lives and the economy. When countries went on lockdown, it disrupted global supply chains, inhibiting international trade and affecting offshoring. This research paper compares the effects of the pandemic on US parent industries with high and low levels of offshoring, specifically by examining datasets on quarterly intermediate inputs and gross output in each industry. I conducted a differences in differences analysis by classifying manufacturing as the treatment group and all other industries with lower offshoring levels as the controlled group. I conclude that higher offshoring industries have seen a greater decrease in productivity, output, and offshoring compared to industries with lower offshoring levels.

1. Introduction

Offshoring has been a controversial topic in economics and politics, where in an increasingly globalized world, businesses have moved operations, and in turn, jobs, abroad. There are many reasons and considerations as to why countries offshore, but the most common ones include: cheaper labor and production in a foreign country, less government regulation such as lower taxes, and an overall comparative advantage. Carluccio et al. (2019) explain that with lower costs of production from international differences in resource prices, companies can produce more efficiently, increasing their profit. In this movement, domestic jobs are often lost or reorganized, so many citizens feel negatively towards offshoring. Ninety-five percent of US respondents oppose businesses' decisions to move manufacturing operations abroad according to Mansfield and Mutz (2013). In the past few years, the coronavirus pandemic has disrupted global supply chains because of country lockdowns, slowing down the spread of the virus and along with it, the international transportation of goods and services. 2020 had the greatest number of industries with decreases in productivity since

2009, because "72 of 90 industries in mining and manufacturing had declines in both output and hours worked" ("72 industries"). Thus, the unprecedented, drastic economic change during the pandemic prompts the question: how have outcomes in US parent companies in high-level and low-level offshoring industries changed before and during the coronavirus pandemic?

Capital moving from one country to another is foreign direct investment (FDI), such as purchasing a factory and hiring workers in another country. This paper focuses on FDI outflow from the US, which occurs when the US purchases or owns at least 10% of a foreign affiliate company. I have used a differences in differences estimation technique to compare manufacturing, the sector with the highest historical offshoring levels, with all other sectors, most notably education services, retail, and mining. I compare the pre-pandemic period of 2019 with the post period starting from the second quarter of 2020 and ending with 2021. The main empirical findings are that there is a significant decrease in intermediate inputs in the manufacturing industry compared to all other industries such as education, retail, and mining, as well as decrease in real gross output in the manufacturing industry compared to retail. The pandemic has had a negative effect on both high- and low-offshore businesses, but because of their high engagement in global value chains, high-offshoring industries have seen a more pronounced decrease in productivity, trade, and offshoring.

This paper is organized as follows. The first section is an introduction and will briefly cover global offshoring patterns. The second section provides background information on offshoring rationale and the coronavirus pandemic's effects on the international economy overall. The third section is a literature review on existing research on offshoring outcomes. The fourth is an overview and description of data used to contrast offshoring outcomes in different industries. The fifth discusses empirical results and theoretical economic models. The sixth section is a conclusion on the main empirical findings and their potential for the future.

2. Background Information

2.1. Offshoring and Global Value Chains

Offshoring should be clearly distinguished as a term individual from outsourcing, which is contracting specific services to a third-party organization either locally or internationally ("The Difference"). Offshoring always occurs internationally and hires complete roles instead of tasks or projects, giving the parent company more direct control over staffing and management, allowing for long term scaling and growth ("The Difference"). Both can coexist as offshore outsourcing, where businesses get services from foreign employees or companies. Background data on industries and priorities used in this paper use these terms interchangeably in cases where they measure offshore outsourcing or similar trends and patterns.

Starting from the early 1980s, global value chains have become more prevalent and crucial to international trade, with technological innovations connecting nations, reducing trade barriers, and the emergence of greater sects of the world participating in a capitalist system, according to Antràs and Chor (2021). Global value chains are, as Antràs and Chor (2021) describe, "a series of

stages involved in producing a product or service that is sold to consumers, with each stage adding value, and with at least two stages being produced in different countries." Intermediate input producers now sell their output to countries all over the world, not just to domestic final good producers, so much so that intermediate input trading makes up two-thirds of world trade, as estimated by Johnson and Noguera (2012). Moreover, Antràs and Chor (2021) conclude that optimal tariffs for final goods are often higher than optimal tariffs for intermediate inputs, indicating a higher demand for imported intermediate inputs in international trade. However, as Lewis et al. (2021) states, this increase in openness, or when trade grows faster than production, makes economies more vulnerable "to changes in trade flows, trade policy, and the composition from trade." Lewis et al. (2021) shows that while more and more of world income is directed from goods to services, this structural change has significantly slowed global trade growth over the forty years from the 1980s to 2020. Without structural change, the global ratio of trade to GDP would be 13% higher than the recorded data.

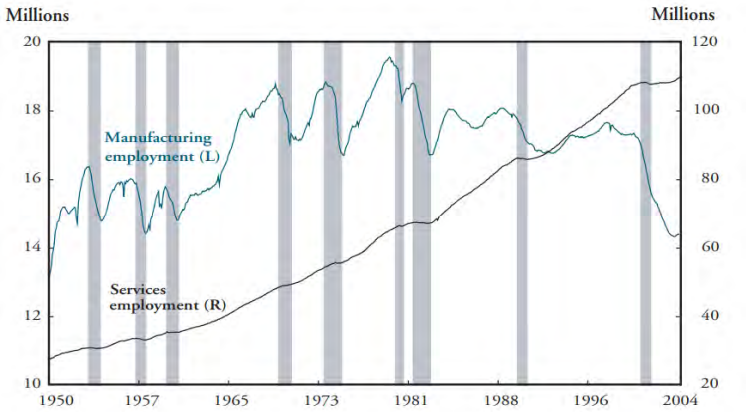


Figure 1. *Jobs Concentrated in Manufacturing in Comparison with Services* (U.S. Department of Labor, Bureau of Labor Statistics, Garner, 2004)

From Figure 1, we can see that manufacturing employment in the US is significantly less than services and is steadily decreasing while services employement remains relatively constant near the end. Garner (2004) explains that manufacturing employment reductions may be due to import competition forcing domestic firms to cut costs and innovate at a faster pace to avoid being displaced. As jobs become increasingly digitized, service jobs are more at risk, specifically information technology careers like telephone call centers moving to, as Garner (2004) states, "low-wage countries, such as India and the Philippines." US parent companies are most likely to offshore jobs that are labor-intensive to other countries to save labor costs and keep capital-intensive jobs, aligning with the Heckscher-Ohlin theorem. Garner (2004) further explains that jobs that are information-based, codifiable, and high-transparency are easier to manage and track remotely, especially given advances in online communication. Thus, location is not as important of a factor, so the job is more prone to offshoring. This

aligns with the conclusion from Blinder (2009) which rates industries and jobs based on "offshorability," or in other words, the probability or likelihood of it being offshored. They concluded that jobs that require employees to be physically present are less likely to offshore.

2.2. Pandemic Changes to Offshoring Objectives

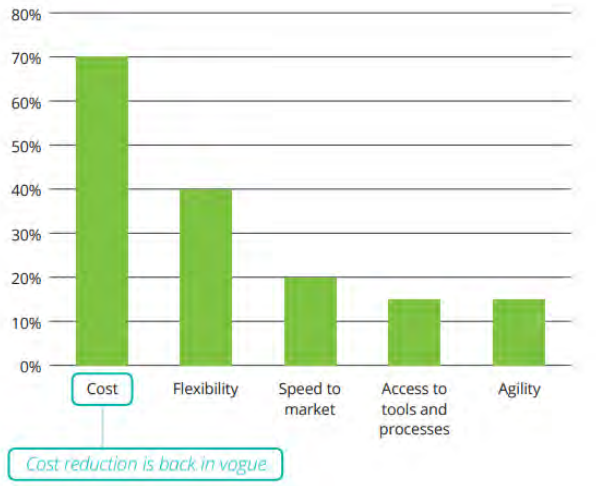


Figure 2. *Offshoring Objectives in Descending Priority Level* (Stoler & Underwood, 2020)

Due to the pandemic imposing quarantines and slowing consumer shopping, there has been a decrease in discretionary income and demand for products, so businesses have made less profit. Thus, as seen in Figure 2, more businesses have prioritized operating cost as the number one motivation for offshoring to balance budgets and ensure there is more income left to sustain business processes, the core business value that remains crucial in times of such uncertainty. To define the factors in Figure 2, we can look to Abdelilah et al. (2018), which describes flexibility as the ability of a foreign affiliate to make changes within the current system when a predicted event happens, whereas agility, although similar, represents altering the overall system in response to an unpredictable event. Speed to market refers to the time it takes for a new product to be developed then produced, and access to tools and processes refers to the firm's role and connections to other firms in the global value chain. Furthermore, COVID has taken a toll on transportation efficiency with more safety checkpoints and understaffed agencies, causing severe delays and backups in global supply chains. Stoler and Underwood (2020) find that from 2018-2020, the trends of having interchangeable providers and multiple offshore firms to fall back on and increased digitization to accommodate new innovations have become more prevalent and have continuously risen. Businesses have identified change management as invaluable characteristics in reacting to unexpected effects brought on by the pandemic and thus are doubling down on these aspects moving forward.

3. Motivation for the Economic Models I Plan on Estimating

I used data from the U.S Bureau of Economic Analysis to plot outcomes over time from 2016-2021 and compare differences between employees, outputs, and intermediate inputs. I interpreted the highest numbers in direct investment based on income as the industries with the highest amount of offshoring. The top two offshoring industries were manufacturing and finance and insurance, and the bottom two were wholesale trade and mining. This industry order is consistent with the estimates in Kajjumba et al. (2020), with manufacturing first and finance second as the industries with the most offshoring, as seen in Figure 3.

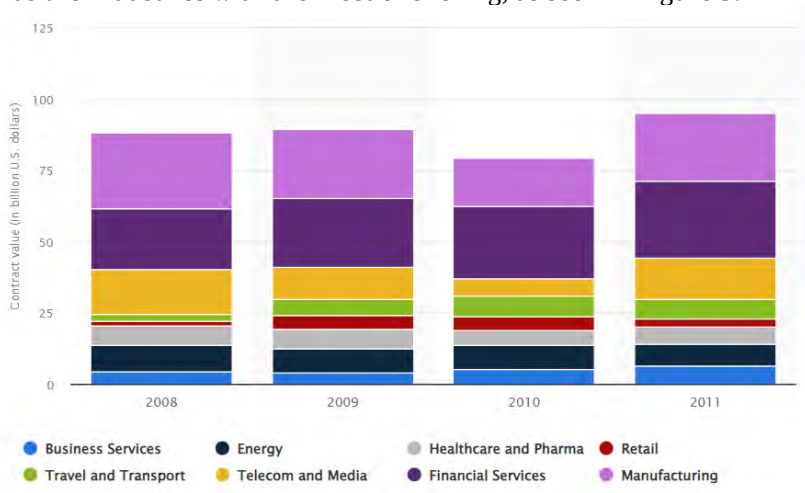


Figure 3. *Comparison of the total contract value in the global outsourcing market by industry from 2008 to 2011 in billion USD* (Kajjumba et al., 2020)

In addition, the overall literature consensus is that offshoring increases firm productivity. According to Couture et al. (2015), a firm that engaged in offshoring's labor productivity was 6.8% higher than that of a similar firm that did not offshore. Through an analysis of intermediate input data, where companies use goods made elsewhere in their final product, Couture et al. (2015) concluded offshoring increases productivity; this is in line with later studies like Feenstra (2016) that also used imported intermediate input data and reallocates labor from less to more productive firms. The starting point for the analysis in this paper is the past analysis of intermediate input data. Furthermore, Hummels et al. (2014) use a dataset of all workers and firms in Denmark to examine the effects of outsourcing on wages. Their main conclusions were that outsourcing lowers wages, specifically that outsourcing to poorer countries lowers wages of less skilled workers more. The type of job also matters in that there is a smaller effect on higher skill critical thinking jobs than trunk strength jobs. This paper's purpose is to validate and verify existing conclusions on offshoring's effect on output by extending its reach to the most recent years and the pandemic, as well as to create new conclusions on unexplored outcomes.

4. Data

In describing the data, this paper addresses the specific outcomes I graphed and performed regression analyses on as well as how the main empirical sources, the Bureau of Economic Analysis (BEA), collects its data. Next, I graphed the annual outcomes from 2016-2020 including wages per full time employee, number of full-time employees, and direct investment income without current cost adjustment. Then, I graphed the trends in 2019-2020 quarterly industry data which had more data points, making it easier to spot larger scale trends more accurately. However, the limitation was that there was only quarterly data on real gross output and real intermediate inputs, so existing annual data may not accurately represent the nuances of change each industry faced.

In order to ensure the accuracy and legitimacy of its data, the BEA conducts seven types of mandatory surveys to collect information on outward and inward direct investment, with time periods ranging between quarterly, annual, and benchmark surveys. Quarterly and annual surveys are different from benchmark surveys in that they “are largely cutoff sample surveys of U.S. parents and their foreign affiliates and of U.S. affiliates of foreign parents above size-exemption levels" ("A Guide" 3). Different measures have their own nuances. For example, "direct investment is recorded in the international investment position accounts at market value, and supplemental information is provided both at historical cost and at current cost," ("A Guide" 1).

Gross output is the sum of the value of the industry’s sales, operating income, inventory change, and commodity taxed that are all valued at purchasers’ prices. Purchasers’ prices are industry paid prices including trade margins, transportation, and excise and sales taxes. In GDP calculations, there is no double counting since only the final output is counted, but gross output does double count since it does not exclude output that may be consumed as intermediate input in the same year. Thus, this measure of gross output may be significantly greater than GDP. Intermediate inputs do not include labor and capital and are also valued at purchasers’ prices. All values in the dataset are listed as the prices in the period they were being observed.

4.1. Descriptive Estimates

The descriptive statistics of mean, median, mode, standard deviation, and range were measured for each of the industries in quarterly real output and quarterly real intermediate input data. Six industries were standardized and compared with each other: manufacturing, finance, wholesale trade, retail trade, mining, and education services. These industries were classified by two-digit NAICS (North American Industry Classification System), the standard codes used by federal agencies to organize statistical data (“Economic Census”). I narrowed the scope to two digits to identify them as overall sectors, so any subsectors of more than three-digits fell under each sector and were added to form a total estimate of the variable in question. Manufacturing had the greatest values while education had the least across all values of descriptive statistics in both input and output data. With an input standard deviation of 138.0299495 and a range of 498.7, manufacturing is the most elastic and volatile compared to education, which has

a standard deviation of 5.382034 and a range of 18. There is a direct relationship between levels of offshoring and volatility since manufacturing, as the industry with the highest offshoring levels, has data that is the most spread out compared to all other industries. This indicates that a reliance on international trade may lead to more frequent and intense fluctuations due to uncontrollable factors in the global economy with many actors and negotiations compared to a more homogenous domestic economy. Similar patterns can be observed in annual data on employees in these respective industries, as indicated in Figures 4 and 5.

Figure 4. *Full-time employees in the manufacturing industry from 2016-2020* ("Full-time Equivalent Employees by Industry.")

Figure 5. *Full-time employees in thousands in the mining, education, retail, and finance industry from 2016-2020* ("Full-time Equivalent Employees by Industry.")

There was a drastic decrease in employment in 2020 during the COVID pandemic in manufacturing and education compared to the slight dip in employment in the mining industry. Manufacturing and mining both have higher

amounts of manual physical labor required in their work compared to education and retail, which both rely on critical thinking and communication more. Due to these unique characteristics, it is significantly easier for the services industries of education and retail to transition to an online business model to continue revenue flows than manufacturing or mining, in compliance with coronavirus regulations. It is more difficult for manufacturing or mining to avoid production shutdowns with their reliance on physical labor and thus cannot adapt to coronavirus restrictions, thereby losing significant revenue. Thus, we can conclude that the higher the offshoring level in an industry, the more susceptible jobs in that industry are to changes in international supply chain operations.

5. Estimation Strategy and Empirical Results

5.1. Theoretical Analysis

The Ricardian model factors in differences in technology as the main motivation for trade and differences in comparative advantage, while the Hecksher-Ohlin model factors in differences in resources such as labor and capital in the absence of a technological difference. Out of two potential export goods, the Heckscher-Ohlin model predicts that each country in a trade agreement will export the good that uses its abundant factor of production and import the other good. However, the model relies on a few crucial assumptions to work, including that final outputs are traded freely but labor and capital do not move between countries.

Offshoring is uniquely different from the Ricardian and Heckscher-Ohlin models because, as Feenstra and Taylor (2017) explain, it does not necessarily trade in final goods. Each good in offshoring can be an intermediate input, produced in different stages in foreign countries and imported to the home country for final assembly. Feenstra and Taylor (2017) find that an increase in offshoring will increase the relative demand and wage for skilled labor in both the foreign and home countries. This conclusion contradicts the Heckscher-Ohlin theorem, which predicts that when two countries open up to trade, if the home country is skilled labor-abundant and the foreign country is unskilled labor-abundant, then the wages of skilled workers relative to the wages of unskilled workers increase in the home country but fall in the foreign country. Carluccio et al. (2019) attribute most of the observed increase in the domestic skill intensity of firms importing from labor-abundant countries to increased offshoring from parent companies. The best option is to use a productions possibility frontier to predict changes in output caused by offshoring. However, in explaining why theoretical models may not align with empirical data, Eaton and Kortum (2001) state that capital goods are heterogeneous, so textbook models are unable to ascertain why countries purchase capital goods from many different sources.

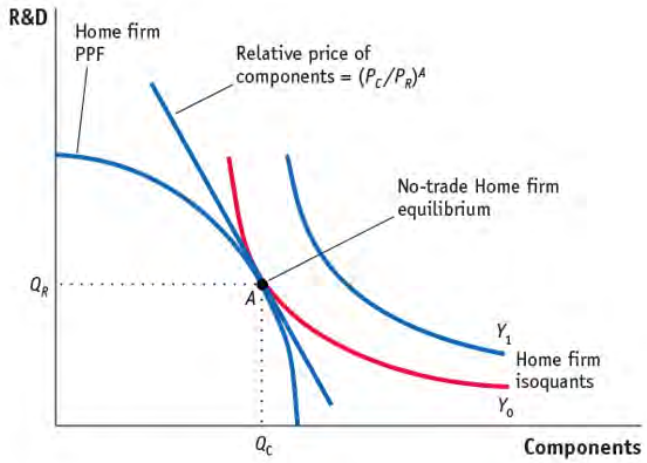


Figure 6. *Production Possibilities Frontier for a firm at no-trade equilibrium* (Feenstra & Taylor, 2017)

We can use the theoretical model of a production possibilities frontier (PPF) to measure gains from offshoring. A PPF displays every possible output combination under the conditions of available resources and technology, comparing the trade-off of producing different goods. As shown in Figure 6, without offshoring, a firm will produce at A with quantities Q_c and Q_r of components and R&D at output level of Y_0. The isoquant or indifference curve Y_0 indicates points along the curve where a firm is producing at constant output but changing inputs in order to determine how much of a final good is produced. Shifting point A to the left indicates that high and low skilled labor will move from production of components into research and development. In Figure 7, since the firm is now open to trade, it can export R&D and import components to start at point B and move along the relative price line, off its production possibilities frontier, to point C. The maximum output amount of Y_1 is produced, an increase from the amount of Y_0 produced without offshoring.

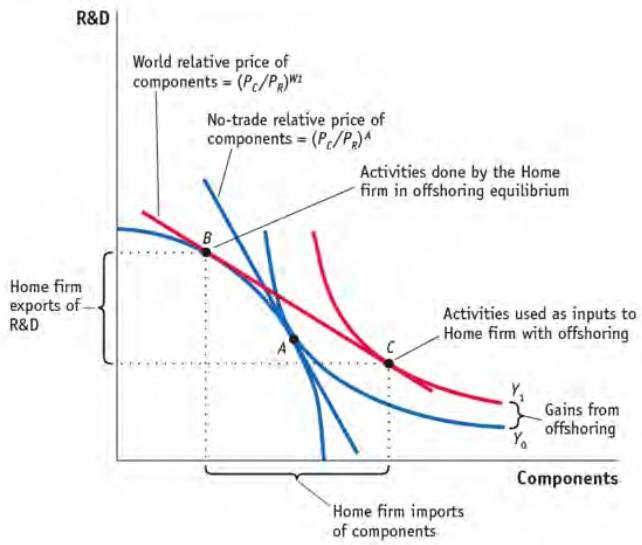


Figure 7. *Production Possibilities Frontier with offshoring at equilibrium* (Feenstra & Taylor, 2017)

We can see that an increase in intermediate inputs is correlated with a rise in imported intermediate inputs, and thus offshoring. When applied to the COVID-19 pandemic, due to lower intermediate inputs and thus, less offshoring from trade blockades, theoretical models predict a loss of previous gains from offshoring where Y_1 will shift down to be closer to Y_0 and point B will shift to the right to be closer to point A. Thus, there is expected to be a decrease in productivity and output in the US during the pandemic compared to pre-COVID-19 years.

5.2. Empirical Estimates

I used a difference in differences estimation strategy to compare the changes in the variables of input and output between the treatment and control groups in the two periods before and during the pandemic. My treatment group consisted of the manufacturing industry because of its high offshoring levels, and the control group consisted of all other industries. The treatment in question is the onset of the COVID-19 pandemic. The dummy variable is y20 for whether the year is during or after the second quarter of 2020 when the pandemic went into full effect.
The following are the econometric models:

$$y = \beta_0 + \delta_0 y20 + \beta_1 manu + \delta_1 y20 \times manu + \mu$$

- β_0 represents average input in 2019 before the pandemic.
- δ_0 indicates changes in the dollar amount of inputs between 2019 and the second quarter of 2020.
- β_1 represents the differences in input in manufacturing and other

industries in the 2020-2021 period based on the industries themselves not due to the impact of the pandemic.

- δ_1 indicates the additional differences in intermediate input in high and low offshoring industries in the 2020-2021 period relative to the 2019 period, the main differences in differences estimator. This estimator can be represented by the equation below.

$$\delta_1 = (\bar{y}_{20,manu} - \bar{y}_{20,oth}) - (\bar{y}_{19,manu} - \bar{y}_{19,oth})$$

The regression analysis with the aforementioned models yields the following outcomes:

Table 1. *Differences in Differences Analysis on Retail Quarterly Output with Manufacturing Industry. (Based on author's own calculations using data from "Real Intermediate Inputs by Industry.")*

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	1803.86	58.96111	30.59406	2.86E-18	1680.869	1926.851
Treated	4159.48	83.3836	49.88367	1.84E-22	3985.545	4333.415
Post	85.28286	77.19822	1.104726	0.282392	-75.7498	246.3155
Treated*Post	-271.166	109.1748	-2.48378	0.021981	-498.9	-43.4311

The only industry quarterly output that produced a statistically significant result is retail when compared to manufacturing. The t statistic is roughly -2.4837, so δ_1 is statistically significant. In analyzing the coefficients, we can conclude that relative to the retail trade industry, manufacturing output went down by 271.166 billion during the pandemic. Thus, manufacturing has lower productivity, indicating a decrease in offshoring production from foreign affiliates. Retail trade is more reliant on selling goods directly to consumers and thus less likely to be offshored since it is advantageous to be closer to the home country and market. The P-value is 0.021, which is less than 0.05, so we can reject the null hypothesis.

Table 2. *Differences in Differences Analysis on Quarterly Intermediate Inputs: Retail vs. Manufacturing Industry (in billions). (Based on author's own calculations using data from "Real Intermediate Inputs by Industry.")*

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	682.56	31.17536674	21.89420916	1.91E-15	617.5293245	747.5906755
Treated	3022.04	44.08862646	68.54466203	3.31E-25	2930.072737	3114.007263
Post	103.6542857	40.81813655	2.539417389	0.019515216	18.5091449	188.7994265
Treated*Post	-317.74	57.72556229	-5.5043205	2.18E-05	-438.15341	-197.326587

Table 3. *Differences in Differences Analysis on Quarterly Intermediate Inputs: Mining vs. Manufacturing Industry (in billions).(Based on author's own calculations using data from "Real Intermediate Inputs by Industry.")*

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	274.16	27.8386	9.848196	4.09E-09	216.0897	332.2303
Treated	3430.44	39.36973	87.13395	2.76E-27	3348.316	3512.564
Post	-55.2886	36.44929	-1.51686	0.144949	-131.32	20.74331
Treated*Post	-158.797	51.54707	-3.08062	0.005898	-266.322	-51.2718

Table 4. *Differences in Differences Analysis on Quarterly Intermediate Inputs: Education vs. Manufacturing Industry (in billions).* (*Based on author's own calculations using data from "Real Intermediate Inputs by Industry.*")

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	99.2	27.5948	3.594881	0.00181	41.63826	156.7617
Treated	3605.4	39.02493	92.38709	8.60E-28	3523.995	3686.805
Post	-4.81429	36.13007	-0.13325	0.895328	-80.1803	70.55172
Treated*Post	-209.271	51.09563	-4.09568	0.000562	-315.855	-102.688

Tables 2 through 4 represent the effects of the pandemic across three industries in comparison to manufacturing; these were chosen for their expected lower levels of offshoring in contrast to wholesale trade and finance, which had more similar offshoring levels to manufacturing. Since the t statistic is -5.504, -3.0806, and -4.095 for retail, mining, and education respectively, δ_1 is statistically significant for all three comparisons. Given the p-value for Treated*Post is significantly less than 0.05, the null hypothesis most likely does not hold true. In analyzing the coefficients, we can conclude that:

- Relative to the retail trade industry, manufacturing intermediate inputs went down by 317.74 billion during the pandemic.
- Relative to the educational services industry, manufacturing intermediate inputs went down by 209.271 billion during the pandemic.
- Relative to the mining industry, manufacturing intermediate inputs went down by 158.797 billion during the pandemic.

The order of industries from greatest to smallest decrease in intermediate inputs is manufacturing, mining, education, and retail. A possible explanation for this is that manufacturing is at the beginning or middle of the supply chain and is more likely to participate in international trade, providing its products to foreign businesses to make up a final good in that foreign market. Retail is the most reliant on direct consumer selling at the end of the supply chain, so it is dealing with final goods with more value added on to the end price before it gets to consumers. Mining and education both follow their respective contrasting trends, with the mining industry focused on quantity of goods produced. Education is a service industry that has more direct teacher and student relationships, so education and retail have the most human interaction compared to all other industries. Industries in the US with higher offshoring levels are most negatively

affected by the pandemic because of their reliance and high participation in global value chains.

We can look to past literature to examine the conclusions and provide supporting evidence of how intermediate inputs affect offshoring. Eaton and Kortum (2001) conclude that the use of manufacturing output in equipment-producing industries is more likely to be used for investment instead of consumption, with roughly half the equipment-producing industries' output used as intermediate inputs. This study relates intermediate inputs to outputs, which can be used to measure productivity. In terms of factor use, Eaton and Kortum (2001) state that equipment-producing industries tend to be more labor and skill intensive when compared to other manufacturing industries. In addition, Boehm et al. (2019) state that a decrease in Japanese inputs caused both a loss of output and a proportionate decrease in imported input use in foreign firms, amplifying the domestic shock in reaction to the 2011 Tohoku earthquake and tsunami internationally. This study provides evidence that foreign affiliates will face additional indirect impacts caused by US intermediate input decreases, such as further output loss on top of their own domestic losses.

Moreover, Elkridge and Harper (2010) find that growth in domestic intermediate inputs directly influence labor productivity, along with growth in capital inputs and technical change. In addition, Amiti and Wei (2005) observe the offshoring of the manufacturing sector through eight years of data from 1992 to 2000 and found that the offshoring of service activities and material inputs added 15-20% of overall productivity growth in the manufacturing industry. This offshoring led to gains for U.S consumers and firms, with lower production costs and thus lower prices. Since both Elkridge and Harper (2010) and Amiti and Wei (2005) find a positive correlation between productivity and offshoring, these conclusions are consistent with the predicted decrease in productivity and offshoring due to the pandemic from the PPF graphs in theoretical analysis.

Thus, we can reasonably conclude that a greater decrease in intermediate inputs in the manufacturing industry compared to other industries shows that the pandemic has a greater effect in decreasing international trade and interrupting global value chains in industries with high offshoring levels compared to industries with lower offshoring levels. These industries with higher offshoring play a large role in global value chains, so the pandemic has disproportionately decreased offshoring and trade levels. This change led to output loss and thus decreased firm productivity, which impacted foreign affiliates and indirectly caused international job loss and economic harm.

6. Conclusion

Immense disruptions to global supply chains have made it more difficult for offshoring operations to continue normally during the COVID-19 pandemic. To prevent the spread of disease and further contamination, many countries have closed their borders, restricting the flow of international trade and investment. Labor supply has diminished as well, with people quarantining and unable to work unskilled jobs that demand physical labor, especially in industries like manufacturing. When offshoring occurs, businesses move capital investments to foreign countries, often to build inputs or parts of the final product to then

assemble and sell domestically. This makes it a controversial topic due to the loss or relocation of domestic jobs and employee income, but most literature points to offshoring increasing firm productivity.

This paper answers the question: how have outcomes in US parent companies in high-level and low-level offshoring industries changed before and during the coronavirus pandemic? In using a differences in differences estimation strategy, I compare the intermediate inputs, gross outputs, wages, and employees of manufacturing, an industry with the highest historical offshoring levels, with all other industries, most notably education services, retail, and mining. The post period was the second quarter of 2020 to 2021 in comparison to 2019, before the pandemic. I conclude there was a decrease in real intermediate inputs in the manufacturing industry compared to other industries, including education, retail, and mining, as well as a decrease in real gross output in the manufacturing industry compared to retail. Overall, the pandemic has had an adverse impact on both high and low offshoring industries, but a more pronounced decline in productivity in high offshoring, because of their high participation in global value chains, leading to lower offshoring levels.

This research advances intermediate input research in the field of offshoring and international trade because it connects a key example of global value chain disruptions to specific industries using the most recent statistics. This can be applied to other recessions to test and further validate conclusions in this paper and to future policies as well to determine which industries may need the most aid or subsidies in times of economic uncertainty. Conclusions from this paper can guide future preparations and reinforcements against unpredictable disruptions to mitigate negative impacts to firms. Future research can focus on offshoring outcomes on foreign affiliates instead of US parents in response to economic interruptions like the coronavirus pandemic and predict how offshoring outcomes will change for the future in the COVID-19 recovery period. Furthermore, more research can be done on how improvements in global communications technology will affect offshoring.

7. References

Abdelilah, B., Bouchra, et al. (2018). Flexibility and Agility: Evolution and Relationship. *Journal of Manufacturing Technology Management*, 29(7), 1138-1162. https://doi.org/10.1108/jmtm-03-2018-0090

Amiti, M., & Wei, S. J. (2005). Service Offshoring, Productivity, and Employment: Evidence from the United States. International Monetary Fund.

Antràs, P., & Chor, D. (2021). Global Value Chains. National Bureau of Economic Research. https://doi:10.3386/w28549.

Aswhite Global. (2022, May 12). The Difference between Offshoring and Outsourcing. https://aswhiteglobal.com/the-difference-between-offshoring-and-outsourcing/

Blinder, A. S. (2009). How Many U.S. Jobs Might be Offshorable? *World Economics*, 10(2), 41-78. https://www.princeton.edu/~ceps/workingpapers/142blinder.pdf

Boehm, C. E., et al. (2019). Input Linkages and the Transmission of Shocks: Firm-Level Evidence from the 2011 Tōhoku Earthquake. *The Review of Economics and Statistics,* 101(1), 60-75. https://doi.org/10.1162/rest_a_00750.

Bureau of Economic Analysis. (2022). Full-time Equivalent Employees by Industry. apps.bea.gov/iTable/iTable.cfm?reqid=19&step=2&isuri=1&1921=survey

Bureau of Economic Analysis. (2022, March 30). Real Gross Intermediate Inputs by Industry. https://www.bls.gov/news.release/pdf/prod2.pdf

Bureau of Economic Analysis. (2022, March 2). Real Gross Output by Industry. https://www.bea.gov/data/industries/gross-output-by-industry

Bureau of Economic Analysis, US Department of Commerce Economics and Statistics Administration. (2018). A Guide to BEA Direct Investment Surveys. apps.bea.gov/survey

Carluccio, et al. (2019). Offshoring and skill-upgrading in French manufacturing. *Journal of International Economics,* 118, 138-159. https://www.sciencedirect.com/science/article/pii/S0022199619300029

Deloitte. (2020, November 19). 2020 Global Outsourcing Survey. https://www2.deloitte.com/global/en/pages/operations/articles/gx-global-outsourcing-survey.html

Eaton, J., & Kortum, S. (2001). Trade in Capital Goods. *European Economic Review,* 45(7), 1195-1235. https://www.sciencedirect.com/science/article/pii/S0014292100001033

Eldridge, L. P., & Harper, M. J. (2010). Effects of Imported Intermediate Inputs on Productivity. US Bureau of Labor Statistics, Monthly Labor Review. www.bls.gov/opub/mlr/2010/06/art1full.pdf

Feenstra, R. C. (2016). Statistics to Measure Offshoring and Its Impact. International Monetary Fund, University of California, Davis. www.imf.org/external/np/seminars/eng/2016/statsforum/pdf/Feenstra_paper.pdf

Feenstra, R. C., & Taylor, A. M. (2017). *International Trade.* (4th ed.). Worth Publishers.

Garner, C. A. (2004). Offshoring in the service sector: economic impact and policy issues. *Economic Review* [Kansas City], 89(3), 5. https://link.gale.com/apps/doc/A123581980/AONE?u=anon~410aef3&sid=googleScholar&xid=3eb598bc

Hummels, D., Jørgensen, B. E., Mutari, E., & Woods, N. (2014). The Wage Effects of Offshoring: Evidence from Danish Matched Worker-Firm Data. *American Economic Review*, 104(6), 1597. https://doi.org/10.1257/aer.104.6.1597

Johnson, R. C., & Noguera, G. (2012). Accounting for Intermediates: Production Sharing and Trade in Value Added. *Journal of International Economics*, 86(2), 224-236. https://www.sciencedirect.com/science/article/pii/S002219961100122X

Kajjumba, G. W., et al. (2020). Offshoring-Outsourcing and Onshoring Tradeoffs: The Impact of Coronavirus on Global Supply Chain. In IntechOpen. https://www.intechopen.com/chapters/74604

Lewis, L. T., et al. (2021). Structural Change and Global Trade. *Journal of the European Economic Association.*

Mansfield, E. D., & Mutz, D. C. (2013). US versus Them: Mass Attitudes toward Offshore Outsourcing. World Politics, 65(4), 571-608. https://doi.org/10.1017/S0043887113000191

U.S. Bureau of Labor Statistics. (2021, May 18). 72 Industries in Mining and Manufacturing Had Declines in Both Output and Hours Worked in 2020. https://www.bls.gov/opub/ted/2021/72-industries-in-mining-and-manufacturing-had-declines-in-both-output-and-hours-worked-in-2020.htm

United States Census Bureau. (2021). Economic Census: NAICS Codes & Understanding Industry Classification Systems. www.census.gov/programs-surveys/economic-census/guidance/understanding-naics.html

Engineering of 1-Dimensional Photonic Crystals with Improved Efficiency for Thermophotovoltaics

Tianyi Yuan

***Author Background**: Tianyi Yuan grew up in China and currently attends Beijing Etown Academy in Beijing, China. His Pioneer research concentration was in the field of engineering and titled "Engineering Photonic Structures with Multi-Layered Films."*

Abstract

Thermophotovoltaics (TPV) is a promising approach to generating electricity from heat and is of large research interest, its main advantages being high fuel compatibility, high portability, high power density, and low maintenance cost. However, traditional TPV systems are limited by low spectral efficiency (mismatch between the emission and absorption spectra). This paper investigates the design of a 1-D photonic crystal as a selective filter in a TPV system, which increases spectral efficiency. The transfer matrix method (TMM) is used for calculation, and the genetic algorithm is used for optimization. Python code is used to implement these algorithms. In four trials of the genetic algorithm, the best design achieves 86.21% spectral efficiency, which is 109% the efficiency of structures in previous studies. This improvement will make TPV systems more competitive against other energy sources. Future research on fabrication tolerance and the energy density of this design is recommended.

1. Introduction

1.1. Thermophotovoltaic Systems

Thermophotovoltaics (TPV) is the technique to convert thermal energy into electricity (Bauer et al., 2011). Today's TPV systems (see Fig. 1) consist of three components: (i) A heat source that usually consumes fuel to deliver thermal energy to the emitter; (ii) A photon emitter that emits electromagnetic radiation of different wavelengths when heated; (iii) A photovoltaic cell (P-N diode) that converts the energy in photons into electricity through the photoelectric effect.

Modern TPV systems have multiple advantages compared to traditional solar cells: (i) higher energy density, (ii) great portability, (iii) high fuel flexibility (the ability to use fossil fuel, nuclear fuel, municipal solid wastes, etc. (Ferrari et al., 2014)), and (iv) the ability for around-the-clock operation (Daneshvar et al., 2015). TPV systems are highly reliable and have low maintenance cost, since their structures do not involve moving parts.

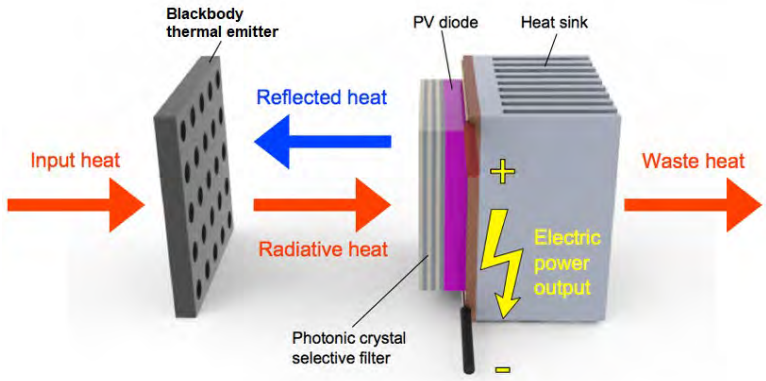


Figure 1. *Schematic diagram of a TPV system. This system uses a blackbody emitter at 1500K and a 1-D photonic crystal selective filter to increase its efficiency (Hernandez, 2018)*

TPV systems have seen wide applications as portable generators (Becker et al., 1999), co-generation systems (Bradbury et al., 2014), combined cycle power plants, solar power plants (Stone et al., 1996), grid connected or independent equipment. TPV systems could also integrate with thermoelectric systems (Qiu et al., 2012) or with Organic Rankine Cycles (De Pascale et al., 2012 and Barbieri et al., 2012) and achieve high system efficiency (Ferrari et al., 2014). Through the development of TPV systems, engineers focus on improving both power output and efficiency (Laroche et al., 2006).

The efficiency of a TPV system relies on multiple factors: (i) The efficiency at which the heat source converts the energy in the fuel to thermal energy, η_{hs}; (ii) the efficiency at which emitted photons fall into the conversion band (wavelength) of the photovoltaic cell, η_{rad}; and (iii) the efficiency at which the P-N diode converts photons into electricity, η_{pn}. The system's overall efficiency is the product of the three efficiencies:

$$\eta_{sys} = \eta_{hs}\eta_{rad}\eta_{pn} \tag{1}$$

A major limitation of TPV systems' efficiency is η_{rad}. For common silicon photovoltaic cells with threshold energy of 1.1eV, the corresponding wavelength is at 1.1μm (see Fig. 2). However, only 6% of the electromagnetic power emitted by a blackbody emitter at $T = 1800K$ can pass this threshold. The radiative efficiency is therefore defined as the energy of photons above the threshold energy of the photovoltaic cell divided by the total energy emitted by the emitter:

$$\eta_{rad} = \frac{P_{\lambda<\lambda_{th}}}{P_{em}} \tag{2}$$

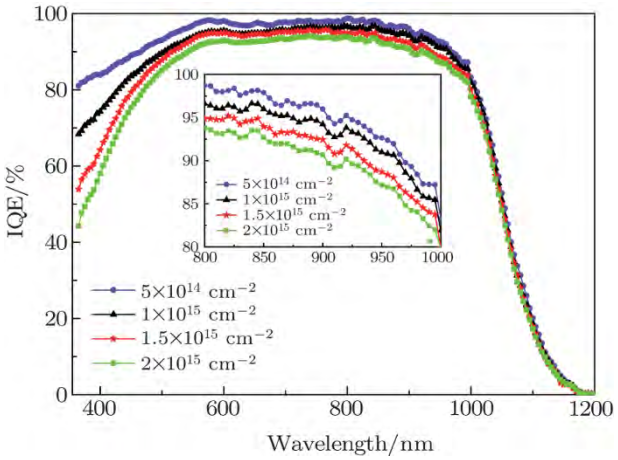


Figure 2. *The conversion efficiency of a silicon PV cell with a band gap wavelength* $\lambda_g = 1.1\mu m$. *The efficiency significantly decreases at longer wavelengths (Peng et al., 2015).*

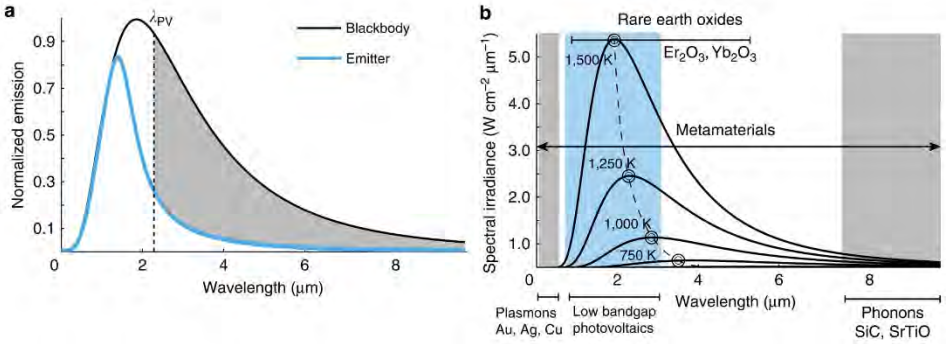


Figure 3. *The radiation spectrum of a* Yb_2O_3 *emitter compared to (a) a black body emitter; (b) bandgap of PV cells (Dyachenko et al., 2016).*

Seeking to improve η_{rad}, engineers have developed different types of spectral control approaches and photon recycling devices. This device reflects photons with below-threshold energy back to the heater and conserves thermal energy. More advanced designs limit the transmission to energy levels slightly above the threshold energy to reduce thermalization losses and further improve efficiency (Rendon-Hernandez, 2018).

Aside from installing filters on the emitter, engineers have created emitters with different materials whose emission spectrum is different from that of a black body. An example is a Yb_2O_3 emitter (see Fig. 3), which emits 20% of its electromagnetic power above the 1.1eV threshold of silicon photovoltaic cells.

1.2. 1-Dimensional Photonic Crystal Filters

Photonic crystals are structures of alternating layers of materials with different refractive indices and optical properties. In 1-Dimensional (1-D) photonic crystals, the repetitive pattern only appears in one direction. 1-D photonic crystals interfere with the propagation of the light, allowing the light of certain wavelengths to propagate without losses and completely blocking (reflecting) light of certain other wavelengths. A continuous wavelength range of complete reflection is called a photonic bandgap (PBG), as shown in Fig. 4A. The photonic bandgap ranges from ~470nm to ~650nm in that particular structure.

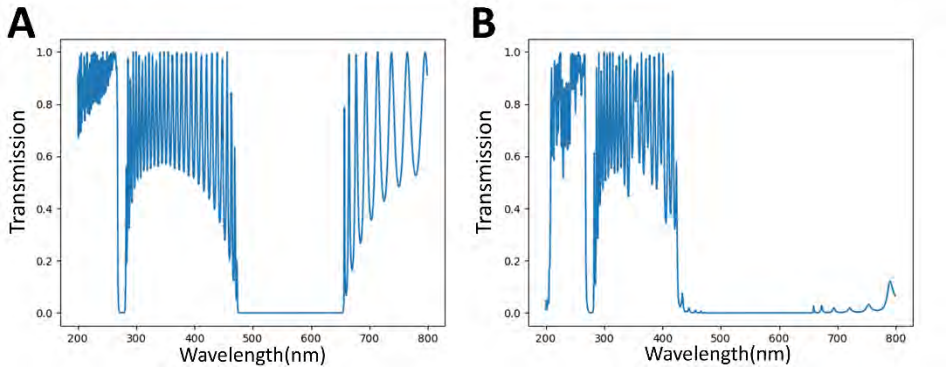


Figure 4. *Transmission diagram of (A) 30 layers of 1-D SiO_2/TiO_2. (B) The same structure with a third material with n = 2.0 inserted in every two pairs of SiO2/TiO2. Three insertions are made.*

The wavelength range of the photonic bandgap is an important quality of photonic crystals. Changing the material, thickness, and periodicity of the photonic crystal can alter its optical properties, thus the photonic bandgap. In practice, engineers also introduce defects to disrupt and modify the photonic bandgap. For example, inserting a third material between every 2 pairs of SiO_2/TiO_2 will extend the photonic bandgap to longer wavelengths, as shown in Fig. 4B.

Magnetic materials can also be introduced into photonic crystals. They interact with light through Magneto-Optic Kerr Effect (MOKE). When a light ray is incident on a magnetized material, it interacts with electrons in that material, and the light's direction of polarization and/or intensity will change (You et al., 1998).

A powerful tool to manipulate the propagation of the light, 1-D photonic crystals have seen wide applications in biosensors (Aly et al., 2020), tunable mirrors and filters (Jena et al., 2019), electro-optic modulators (Pan et al., 2015), refractive index sensor (Nunes et al., 2010), and thermophotovoltaic cells (Celanovic et al., 2004).

2. Methodology

In this paper, Transfer Matrix Method (TMM) is used to predict the performance of a given photonic crystal design. A Python program is developed to carry out the calculations according to TMM and plot the transmission diagram of the given photonic crystal.

2.1 Transfer Matrix Method

Transfer Matrix Method (TMM) is a powerful tool that applies to paraxial rays. It can effectively predict the behavior of light rays propagating in one plane, so TMM is suitable for calculating the case of 1-D photonic crystals. Compared to traditional Plane Wave Expansion Method, TMM converges faster and uses fewer plane waves (Li et al., 2003), which would significantly improve the calculation speed. When the light passes through a cascade of optical materials (as in the case in 1-D photonic crystals), the resulting ray is simply the product of all matrices that represent the materials and the incident ray (Zhan et al., 2013). The controlling idea of TMM is to use a $\mathbf{1} \times \mathbf{2}$ matrix to describe the light ray's position and direction, and use a $\mathbf{2} \times \mathbf{2}$ matrix as an operator to represent the effect of the optical material on the light ray. The relationship of the matrices is shown below:

$$\begin{bmatrix} \boldsymbol{y_2} \\ \boldsymbol{\theta_2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} \begin{bmatrix} \boldsymbol{y_1} \\ \boldsymbol{\theta_1} \end{bmatrix} \tag{3}$$

As shown in Eqn. 3, the $\mathbf{2} \times \mathbf{2}$ matrix is the operator that describes the changes of the position $\boldsymbol{y_1}$ and the direction $\boldsymbol{\theta_1}$ of the incident ray. For 1-D photonic crystals, two matrices are used for TMM: boundary matrix and propagation matrix. The boundary matrix describes the case when the light ray propagates through the interface of two media, 1 and 2:

$$\boldsymbol{B_{12}} = \frac{\mathbf{1}}{\mathbf{2}\boldsymbol{n_2}} \begin{bmatrix} \boldsymbol{n_1} + \boldsymbol{n_2} & \boldsymbol{n_2} - \boldsymbol{n_1} \\ \boldsymbol{n_2} - \boldsymbol{n_1} & \boldsymbol{n_1} + \boldsymbol{n_2} \end{bmatrix} \tag{4}$$

where $\boldsymbol{n_1}$ and $\boldsymbol{n_2}$ are the refractive indices of material 1 and 2 respectively.

The propagation matrix describes the case when the light of wavelength $\boldsymbol{\lambda}$ propagates through a uniform media with refractive index $\boldsymbol{n_1}$ and thickness $\boldsymbol{d_1}$:

$$\boldsymbol{P_1} = \begin{bmatrix} \boldsymbol{e^{-i\phi_1}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{e^{-i\phi_1}} \end{bmatrix} \tag{5}$$

where $\boldsymbol{\phi_1} = \frac{\boldsymbol{2\pi}}{\boldsymbol{\lambda}} \boldsymbol{n_1 d_1}$.

With these matrices defined, we can calculate the behavior of the light as it propagates through any given 1-D photonic crystals. An example is these two layers of SiO_2 and TiO_2, with refractive indices $\boldsymbol{n_1} = \mathbf{1.5}$ and $\boldsymbol{n_2} = \mathbf{2.5}$, thickness $\boldsymbol{d_1} = \mathbf{100}$nm and $\boldsymbol{d_2} = \mathbf{50}$nm respectively. The light propagates through this complex in five steps: (i) Entering SiO2 from the air; (ii) Propagating in SiO2; (iii) Entering TiO_2 from SiO_2; (iv) Propagating in TiO_2; (v) Entering the air from TiO_2. We then reverse the sequence of the five steps. Starting from step (v), we multiply the input matrix with the corresponding transfer matrix of each step until the first step:

$$\begin{bmatrix} \boldsymbol{y_2} \\ \boldsymbol{\theta_2} \end{bmatrix} = \boldsymbol{B_{air1}} \times \left(\boldsymbol{P_1} \times \left(\boldsymbol{B_{12}} \times \left(\boldsymbol{P_2} \times \left(\boldsymbol{B_{2air}} \begin{bmatrix} \boldsymbol{y_1} \\ \boldsymbol{\theta_1} \end{bmatrix} \right) \right) \right) \right) \tag{6}$$

2.2 Genetic Algorithm

The genetic algorithm is a global optimization method based on group genetics (Kumar et al., 2010). It incorporated the mechanics of genetic mutation, natural selection, and crossing over. In the genetic algorithm, each solution is modeled into

a chromosome. The calculation starts with an initial population (solution set). The initial population is then allowed to mate and produce offspring by exchanging and combining the chromosomes of two individuals. A fitness function is introduced to calculate the fitness score of each individual. Individuals with low fitness scores are removed from the population by natural selection, and high-fitness individuals are allowed to mate again and produce the next generation. The evolution continues until the solution satisfies the end condition.

Compared to local optimization methods, the genetic algorithm is more effective in solving problems that do not already have a well-defined efficient solution (Kumar et al., 2010). Therefore, the genetic algorithm is a powerful tool for designing photonic crystals of different purposes, since the theoretical ideal photonic crystal for TPV cannot be achieved in practice.

To design a photonic crystal using the genetic algorithm, we first model the structures of 1-D photonic crystals as chromosomes. Photonic crystals consist of materials with different refractive indices and thicknesses, and each material is denoted by a number. A number string represents the exact structure and material of a photonic crystal.

To give a quantitative evaluation of each structure's fitness, a fitness function is introduced. The function does not calculate the efficiency of the filter directly (for the efficiency of the filter, see section 3.2), but it rather compares the efficiency of the design to that of an ideal filter. An ideal filter is defined as a structure that allows full transmission for photons above the threshold frequency (or below the threshold wavelength) and reflects all photons below the threshold frequency (or above the threshold wavelength):

$$\boldsymbol{T}(\boldsymbol{\lambda}) = \begin{cases} \mathbf{1}, & \boldsymbol{\lambda} < \boldsymbol{\lambda_g} \\ \mathbf{0}, & \boldsymbol{\lambda} \geq \boldsymbol{\lambda_g} \end{cases} \tag{7}$$

where $\boldsymbol{T}(\boldsymbol{\lambda})$ represents the transmission of the filter at wavelength $\boldsymbol{\lambda}$.

Assuming the emitter is a blackbody, we apply Planck's Law to calculate its radiation energy distribution at $\boldsymbol{T} = \mathbf{1800K}$, which is the operational temperature of most TPV systems:

$$\boldsymbol{B_\lambda}(\boldsymbol{\lambda}, \boldsymbol{T}) = \frac{\boldsymbol{2hc^2}}{\boldsymbol{\lambda^5}} \frac{\mathbf{1}}{\boldsymbol{exp}(\frac{\boldsymbol{hc}}{\boldsymbol{\lambda k_B T}}) - \mathbf{1}} \tag{8}$$

For a GaSb photovoltaic cell, the threshold wavelength $\boldsymbol{\lambda_g} = \mathbf{1.78\mu m}$. Therefore, we can calculate the total transmitted power and reflected power of an ideal filter:

$$\boldsymbol{P_{tr}} = \int_{\mathbf{0}}^{\boldsymbol{\lambda_g}} \boldsymbol{B_\lambda}(\boldsymbol{\lambda}, \mathbf{1800}) \boldsymbol{T}(\boldsymbol{\lambda}) \mathrm{d}\boldsymbol{\lambda} \tag{9}$$

$$\boldsymbol{P_{re}} = \int_{\boldsymbol{\lambda_g}}^{\boldsymbol{\infty}} \boldsymbol{B_\lambda}(\boldsymbol{\lambda}, \mathbf{1800}) \boldsymbol{T}(\boldsymbol{\lambda}) \mathrm{d}\boldsymbol{\lambda} \tag{10}$$

Due to limitations of the calculation speed and the transfer matrix method, the transmission function $\boldsymbol{T}(\boldsymbol{\lambda})$ is not available. Instead, the transmission value is calculated for every discrete wavelength. Therefore, eqns. 9-10 are discretized:

$$\boldsymbol{P_{tr}} = \sum_{\boldsymbol{\lambda}=\mathbf{500}}^{\mathbf{1779}} \boldsymbol{B_\lambda}(\boldsymbol{\lambda}, \mathbf{1800}) \boldsymbol{T}(\boldsymbol{\lambda}) \tag{11}$$

$$P_{re} = \sum_{\lambda=1780}^{5000} B_{\lambda}(\lambda, 1800) T(\lambda) \quad (12)$$

The lower and upper bounds, due to the limitation of the discretization method, are restricted to [**500nm**, **5000nm**]. Specifying the wavelength range only slightly affects the precision, since most radiative power is inside this wavelength range. The radiative energy distribution below and above the threshold is uneven. We want to assign the same weight for transmission and reflection. Therefore, P_{tr} and P_{re} are divided by the total radiative power in their respective wavelength range, and the efficiency indices of transmission and reflection are defined:

$$\eta_{tr} = \frac{\sum_{\lambda=500}^{1779} B_{\lambda}(\lambda, 1800) T(\lambda)}{\sum_{\lambda=500}^{1779} B_{\lambda}(\lambda, 1800)} \quad (13)$$

$$\eta_{re} = \frac{\sum_{\lambda=1780}^{5000} B_{\lambda}(\lambda, 1800) T(\lambda)}{\sum_{\lambda=1780}^{5000} B_{\lambda}(\lambda, 1800)} \quad (14)$$

For an ideal filter, $\eta_{tr} = \eta_{re} = 1$. The sum of efficiency indices of an ideal filter is $\eta_{tr} + \eta_{re} = 2$. The fitness function is therefore defined as the sum of the efficiency indices of an ideal filter divided by the sum of efficiency indices of a given photonic crystal design:

$$\text{Fitness} = \frac{2}{\eta_{tr} + \eta_{re}} \quad (15)$$

According to eqn. 15, the photonic crystal becomes closer to ideal (functions as an ideal filter) as the fitness score gets closer to 1.0. The best fitness score possible is 1.0, where the photonic crystal is already an ideal one.

2.3 Python Implementation of the Genetic Algorithm

Python 3.9 is used to perform the calculation of the transfer matrix method and the genetic algorithm. Matplotlib is used to visualize the result (generate plots). Two arrays, $n[i]$ and $d[i]$, store the refractive index and the thickness of material i. The index i is a single gene. Each of SiO_2 and Si has 11 genes of different thickness, whose index is 1 to 22. We define a function, PLAW(LAMBDA), to calculate the blackbody radiative energy at the given wavelength and 1800K. The result is added to another array, BB[Lambda], and is used for the fitness calculation.

The genetic algorithm (see Fig. 5 and Appendix A) starts with an initial population. The initial population consists of individuals whose structure is completely random. First, the program picks an integer from [8,12] randomly as the length (number of layers) of the individual photonic crystal, since most previous works focus on photonic crystals at ~10 layers. Then, the program picks random genes from the gene pool and adds them to this individual. This process repeats until the initial population has 1000 individuals.

For each individual in the initial population, the fitness score is calculated according to the following procedures: the wavelength lambda is looped through 500nm to 5000nm with a 1nm step size. The transmission of the photonic crystal is calculated at this wavelength. When lambda is smaller than 1780nm, the

transmission is multiplied by the blackbody radiative energy, BB[lambda], and this product is added to the transmission score. When lambda is greater than 1780nm, the reflection is multiplied by BB[lambda], and the product is added to the reflection score. When the loop finishes, each transmission and reflection score is divided by the maximum score possible (the transmission/reflection score of an ideal filter). The quotients are added, and their reciprocal is the fitness score.

After each individual has a fitness score, all individuals are sorted in descending order. Individuals in the top 10% fitness are passed on to the next generation. Individuals in the top 50% fitness are allowed to cross over and produce the rest of the individuals of the next generation. Each time a new gene is given to the children, there is a 10% chance for it to mutate, so the gene is replaced by a random gene. The program repeats the fitness calculation and crossing over until reaching the 200th generation or detecting convergence.

Convergence is defined as the program obtaining the same solution in 10 successive generations. In this case, the program is terminated because it may have found the global optimum solution under the given boundary conditions.

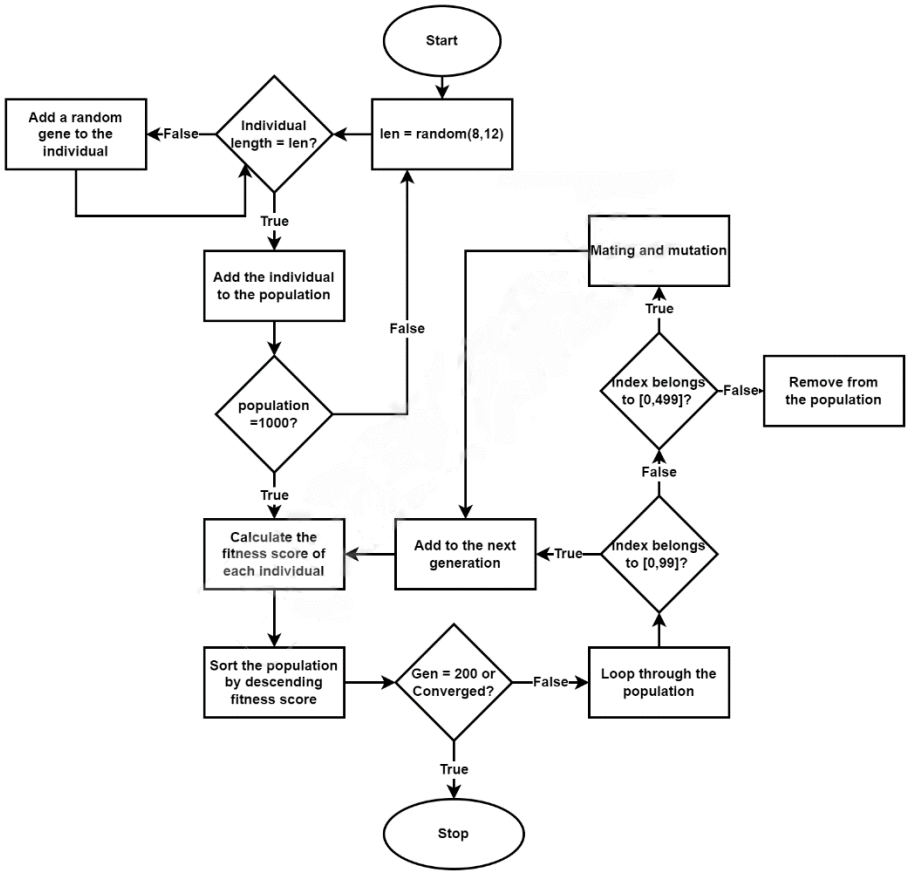


Figure 5. *The flowchart of the Python program.*

3. Modeling of a Thermophotovoltaic System

The thermophotovoltaic system consists of different components. As discussed in section 1.1, the system efficiency depends on the product of all components. This section discusses the efficiency of the system based on Planck's Emission Law. Then, a filter consisting of Si/SiO2 is introduced for further discussion.

3.1 The Structure of 1-D Si/SiO_2 Photonic Crystals

The optical characteristics of Si and SiO_2 are well investigated. Mao et al. have proposed designs of Si/SiO_2 photonic crystals from theoretical analysis. The structure is designed according to the quarter wave theory, which enhances the interference of the beam to achieve high transmittance and reflectance. The quarter wave theory suggests that the central wavelength of the photonic bandgap, $\boldsymbol{\lambda_0}$, is (Mao et al., 2010):

$$\boldsymbol{\lambda_0 = \lambda_1\left(1+\left(\frac{2}{\pi}\sin^{-1}\left(\frac{n_{Si}-n_{SiO_2}}{n_{Si}+n_{SiO_2}}\right)\right)\right)} \quad (16)$$

where $\boldsymbol{\lambda_1}$ is the minimum wavelength (left boundary) of the photonic band gap. According to eqn. 16, we take $\boldsymbol{\lambda_1 = 1.78\mu m}$ to obtain $\boldsymbol{\lambda_0}$, and the thickness of Si and SiO_2 layers can be calculated (Khosroshahi et al., 2017):

$$\boldsymbol{d_{Si} = \frac{\lambda_0}{4n_{Si}}} \quad (17)$$

$$\boldsymbol{d_{SiO_2} = \frac{\lambda_0}{4n_{SiO_2}}} \quad (18)$$

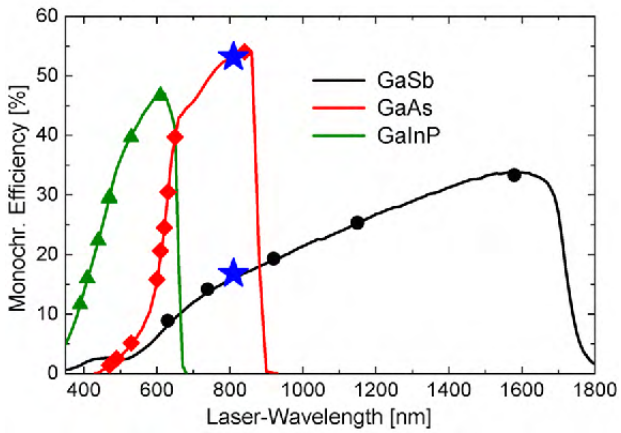


Figure 6. *The conversion efficiency of a GaSb photovoltaic cell (black) at different wavelengths (Bett et al., 2008).*

3.2 The Efficiency of the Emitter-Photovoltaic Cell System

A thermophotovoltaic system of a blackbody emitter at 1800K, a 1-D photonic crystal filter, and a GaSb photovoltaic cell (see Fig. 6, threshold energy $\boldsymbol{E_g = 0.72eV}$, and threshold wavelength $\boldsymbol{\lambda_g = 1.1\mu m}$) are considered. As discussed in section 1.1, the radiative efficiency is defined as the above-bandgap energy of the emitter over the total energy delivered from the emitter to the filter (see eqn. 2). According to the Planck's Law, the energies $\boldsymbol{P_{\lambda>\lambda_{th}}}$ and $\boldsymbol{P_{em}}$ can be calculated by integration:

$$\boldsymbol{P_{\lambda>\lambda_{th}} = \frac{2\pi}{c^2h^3}\int_{E_g}^{\infty} \frac{E^3}{\exp\left(\frac{E}{kT}\right)-1} T_f(E)\,dE} \quad (19)$$

$$\boldsymbol{P_{em} = \frac{2\pi}{c^2h^3}\int_{0}^{\infty} \frac{E^3}{\exp\left(\frac{E}{kT}\right)-1} T_f(E)\,dE} \quad (20)$$

where $\boldsymbol{E_g}$ is the threshold energy of the photovoltaic cell and $\boldsymbol{T_f(E)}$ is the transmittance of the photonic crystal filter for electromagnetic waves with energy $\boldsymbol{E = \frac{hc}{\lambda}}$. For an ideal filter, $\boldsymbol{\eta_{rad} = 1}$ because $\boldsymbol{T(\lambda) = 0}$ for $\boldsymbol{\lambda \geq \lambda_g}$.

4. Results and Discussion

Typical thermophotovoltaic systems with proper spectral control achieve a spectral efficiency at ~75%. For example, Fraas et al. proposed a TPV system with a tungsten emitter at 1525K and a GaSb photovoltaic cell, which attains 75% spectral efficiency (Fraas et al., 2000). A SiC emitter at 1325K coupled to an InGaAs PV cell also achieves 75% spectral efficiency (Fraas et al., 2003). The overall efficiency of these systems is 20% and even higher, and they have an exceptional power density ($\sim \boldsymbol{3 \times 10^4 W/m^2}$) (Fraas et al., 2003). The most efficient type of spectral control is the tandem filter, which achieves a ~76% efficiency when coupled to a 0.6eV low gap photovoltaic cell and ~83% efficiency with a 0.52eV cell (Fourspring et al., 2004). However, tandem photonic crystals are limited by their complex structures and fabrication difficulty. Traditional 1-D photonic crystals, when properly designed and optimized, may achieve the efficiency of the tandem photonic crystals. Mao et al. produced a 1-D photonic crystal design with SiO_2 and Si, which attained ~55% spectral efficiency at T=1800K. The structure is denoted by L/2[LH]^4, the first L/2 being a half-thickness layer of SiO_2. Based on Mao et al.'s design, we improve the theoretical spectral efficiency of the photonic crystal using the genetic algorithm (Mao et al., 2010).

4.1 Genetic Algorithm Setup and Results

The genome base consists of three genes: SiO_2($\boldsymbol{d = 390nm}$), Si($\boldsymbol{d = 170nm}$), and SiO_2 ($\boldsymbol{d = 195nm}$, the half-thickness layer). The population size is 100 and the mutation rate is 10%. In the 7th generation, an improved sequence is generated, which has a fitness score of $\boldsymbol{F = 1.269}$. This design is denoted as LH/2[LH]^{2}L/2, which removes two pairs of Si/SiO_2 and adds another half-thickness SiO_2 layer.

In the second trial, all other conditions remain unchanged, except for

increasing the population size to 1000. In this trial, the structure is further improved by adding one pair of Si/SiO_2 that is removed in trial 1. The new structure, denoted by $LH/2[LH]^3L/2$, has a fitness score of $\mathbf{F = 1.244}$.

In the third trial, the genome pool is enlarged to 10 genes. The genes are Si/SiO2 of different thicknesses. As prescribed by the quarter wave theorem, theoretically, the refractive indices $\boldsymbol{n_1}$ and $\boldsymbol{n_2}$, and thickness $\boldsymbol{d_1}$ and $\boldsymbol{d_2}$, shall satisfy the following relations:

$$\frac{\boldsymbol{n_1 d_1}}{\boldsymbol{n_2 d_2}} = \mathbf{1} \qquad (\mathbf{21})$$

In the third trial, this ratio, $\frac{\boldsymbol{n_1 d_1}}{\boldsymbol{n_2 d_2}}$, is varied to introduce new genes. Corresponding genes satisfying $\frac{\boldsymbol{n_1 d_1}}{\boldsymbol{n_2 d_2}} = \mathbf{0.5, 0.67, 0.75, 1.33, 1.5, and\ 2}$ are added to the gene pool. The results converge slowly and yield the same results as in trial 1. This trial is terminated at generation 18.

In the fourth trial, the thickness of SiO_2 and Si are slightly varied around the thickness specified by the quarter wave theory (see Appendix A). For each material, 11 genes of different thicknesses are added to the gene pool. Convergence is observed at generation 24 with a fitness score of $\mathbf{F = 1.213}$. We name this optimal structure $L/2H[LH]^3L/2$ Mod., and we discuss its optical characteristics in the following sections.

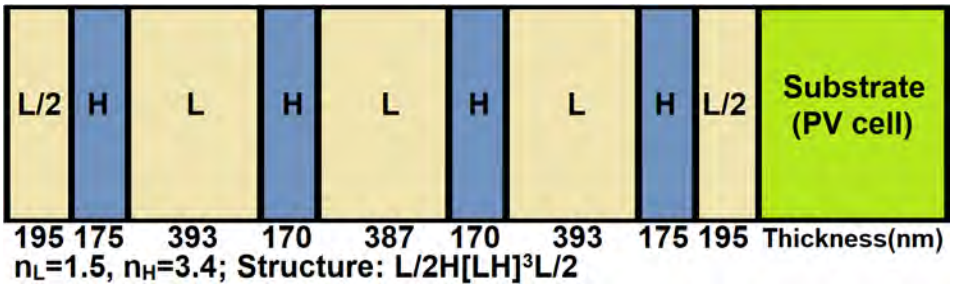


Figure 7. *The schematic illustration of the photonic crystal structure obtained in trial 4.*

According to eqns. 19-20, the efficiency of the photonic crystal filters can be calculated. However, due to the limitation of the numerical method, the integration cannot be performed. Instead, the result is summed discretely at each integral wavelength from 500nm to 5000nm. The efficiency at every integral temperature between [1200K,1800K] is calculated (see Fig. 8A and Table 1). As temperature increases, spectral efficiency increases slightly. The unfiltered efficiency increases almost linearly as the temperature increases (the red line). However, even at 1800K, an unfiltered TPV system only achieves 36% spectral efficiency compared to 86% spectral efficiency of a filtered system.

The final design ($L/2H[LH]^3L/2$ Mod.) significantly improves the pass-band behavior between 880nm and 1760nm while maintaining the photonic band gap's width and position. The original design by Mao et al. (denoted by $L/2[LH]^4$) has a photonic bandgap from 1840nm to 3220nm, and the final design's band gap ranges from 1890nm to 3280nm. After the band gap, two minor transmission peaks

lead to loss of energy. Compared to the original design, the final design shifts these peaks to longer wavelengths, which reduces the loss. However, both designs fail to eliminate the minor reflection band from 710nm to 850nm. Further research is suggested to focus on expanding the major band gap and removing the minor reflection gap.

Table 1. *The efficiency of a TPV system at selected temperatures with different filter designs.*

Temperature	No filter	$L/2H[LH]^4$	$L/2H[LH]^3L/2$	$L/2H[LH]^3L/2$ Mod.
1200K	12.12%	74.24%	77.81%	78.89%
1400K	19.77%	77.24%	81.64%	82.44%
1600K	27.90%	78.87%	84.18%	84.83%
1800K	35.82%	79.47%	85.72%	86.21%

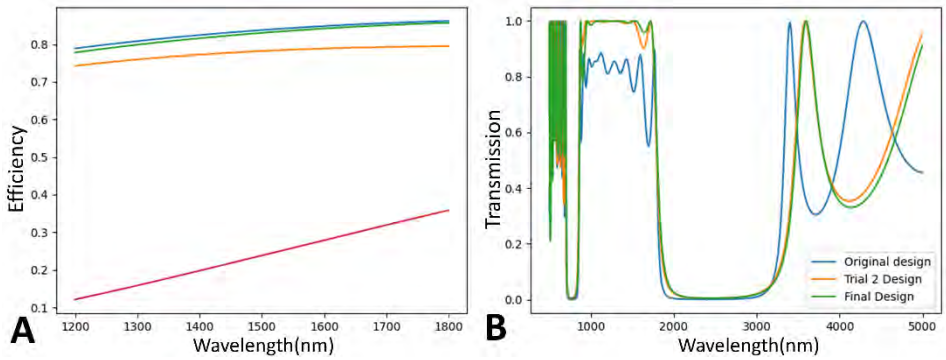


Figure 8. *The (A) spectral efficiency of the TPV system with and without (red) photonic crystal filters: Mao et al.'s design (orange), trial 2 (green), and trial 4 (blue); And their (B) transmission: Mao et al.'s design (blue), trial 2 (orange), and trial 4 (green).*

4.2 Manual Optimization and Verification

Although the genetic algorithm is a powerful optimization tool, it may fail to produce the global optimal solution in rare situations. In this paper, the values of important parameters (mutation and selection rate, fitness function, TMM function) and constants may lead to system error. Therefore, it is important to examine the solution manually and try to modify it for better performance. In four trials, the genetic algorithm strictly limits the length of the design to 8-12 layers (which aims for better converging speed) and does not consider other potential improvements (e.g., an anti-reflection coating). Therefore, we change these parameters manually to see if the design can further improve.

In trial 1, the algorithm removes two pairs of [LH], and then in trial 2, one pair is added back. Since the band gap characteristics are sensitive to the number of repetitive pairs, we want to investigate that our final structure, $L/2H[LH]^3L/2$, is the best arrangement. We start with SiO_2 and Si to investigate this relation.

The repetitive [LH] pairs at the center are varied from 2 to 5 pairs. In the

band gap region, more repetitive pairs lead to a deeper band gap (see Fig. 10A), so increasing the number of pairs improves the reflection band performance. However, increasing the number of pairs hampers the pass band transmission by introducing troughs (see Fig. 10B) to the transmission plateau. The L/2H[LH]^{5}L/2 structure has 1 major trough and 1 minor trough at the right end of the plateau, which significantly compromises the efficiency. This trade-off between pass band performance and band gap performance can be further studied, so engineers may find the balance point. For thermophotovoltaic cells, both L/2[LH]^{3}L/2 and L/2[LH]^{4}L/2 structures exhibit good band gap and pass band performance.

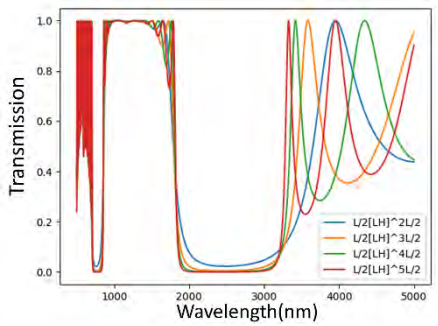


Figure 9. *The transmission of photonic crystal designs with different numbers of [LH] pairs.*

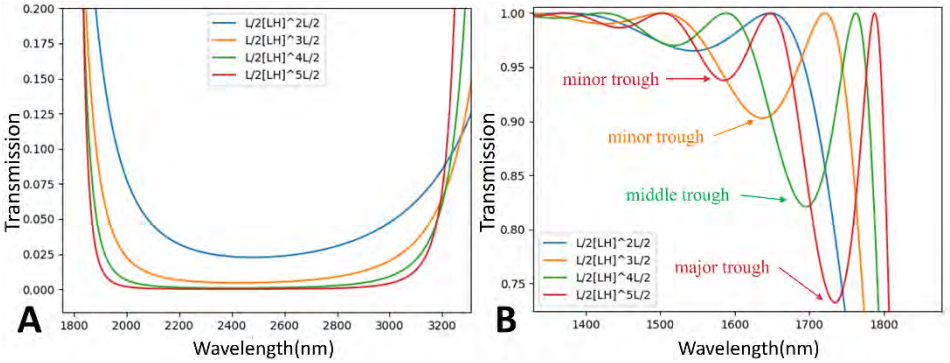


Figure 10. *The (A) band gap and (B) pass band characteristics of photonic crystals with different numbers of [LH] pairs.*

To further increase the pass band efficiency, an anti-reflection (AR) coating is introduced (Menna et al., 2015). An AR coating is a thin film of material whose refractive index satisfies (Brown et al., 2004):

$$n_{AR} = \sqrt{n_1 n_2} \tag{22}$$

where n_0 and n_1 are refractive indices of the layers besides the AR coating. In this paper, the AR coating is installed between air and the first L/2 layer, so n_0 is 1.0 and n_1 is 1.5. AR coatings of different thicknesses are included as the first layer of the structure, but none of these altered structures produce better results. As shown in

Fig. 11, the AR coating reduces the efficiency of the pass band and does not remove the minor band gap at ~700nm.

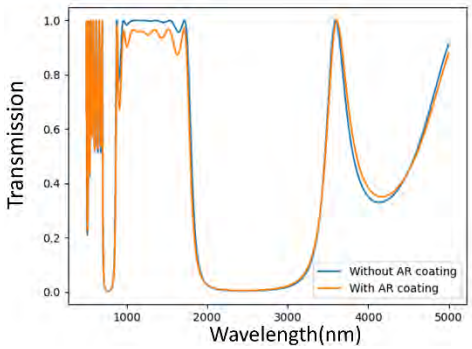


Figure 11. *The transmission of photonic crystal designs with and without an AR coating at the first layer.*

4.3 Summary and Conclusion

This paper implements the genetic algorithm to research the best structure of a Si/SiO_2 photonic crystal for a thermophotovoltaic system. Compared to previous designs, the final design in this paper improves the spectral efficiency from 79% to 86%. This result will make photonic crystal filters a strong candidate for spectral control in TPV systems. The genetic algorithm in this study turns out to be a versatile tool in engineering. Changing the parameters of the algorithm will allow it to find solutions for different design purposes.

This paper, however, does not cover all characteristics of the final design ($L/2H[LH]^3L/2H$ Mod.). Further research on the behavior of this design and its fabrication tolerance is encouraged. Researchers may further improve the efficiency of this design by using different crossing over, mutation, and natural selection methods for the genetic algorithm. Moreover, installing any type of selective filter reduces the power density of the thermophotovoltaic system. Engineers are encouraged to consider this trade-off between the power density and efficiency (Pirvaram et al., 2021) when choosing a filter for a system.

Appendix A

Python code for the genetic algorithm:

```
import numpy as np
import math
import random
import matplotlib.pyplot as plt
h = 6.626e-34 #physical constants
kb = 1.3806e-23
c = 2.9979e8
idealT = 6030.190919887218
idealR = 12542.754861018404
def PLAW(LAMBDA): #blackbody radiation function
    LAMBDA = LAMBDA * 1e-9
    return 2*h*c**2/(LAMBDA**5)/(np.exp(h*c/LAMBDA/kb/1800)-1)
POPUL_Size = 1000 #population size
BB = [] #stores the power of black body radiation
for i in range(500,5000):
    BB.append(PLAW(i)/1e10)
n = [1.0, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 3.4, 3.4, 3.4, 3.4, 3.4, 3.4, 3.4, 3.4, 3.4, 3.4, 3.4, 3.4, 3.4, 3.4, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 1.5, 3.4, 3.4]
d = [114, 390, 389, 388, 391, 392, 195, 194, 193, 196, 197, 393, 387, 170, 171, 172, 173, 174, 175, 169, 168, 167, 166, 165, 394, 395, 386, 385, 192, 191, 190, 198, 199, 200, 176, 177]
Gen_Mat = list(range(1,len(n)-1))
for i in Gen_Mat:
    i = str(i)
def boundary_matrix(N1,N2): #returns to the boundary matrix
    return 1/2/N2*np.array([[N2+N1,N2-N1],[N2-N1,N2+N1]])
def propagation_matrix(Phi): #returns to the propagation matrix
    return np.array([[np.exp(-1j*Phi),0],[0,np.exp(1j*Phi)]])
def phaseangle(media,wavelength): #calculates the phase angle
    return n[media]*d[media]*2*np.pi/wavelength
class Individual(object):
    def __init__(self, chromosome): #initialize the field variables
        self.chromosome = chromosome
        self.fitness = self.cal_fitness()
    @classmethod
    def mutated_genes(self):#returns to a random gene in the gene pool
        global Gen_Mat
        gene = random.choice(Gen_Mat)
        return gene
    @classmethod
    def create_gnome(self):#creates the genome of an individual
        gnome_len = 8 + random.randint(0,4)#randomize the length within [8,14]
        return [self.mutated_genes() for _ in range(gnome_len)] #randomly choose the initial genes
    def mate(self, par2):#takes in two individuals, returns a child chromosome
        child_chromosome = []
        for gp1, gp2 in zip(self.chromosome, par2.chromosome):
            prob = random.random()
            if prob < 0.45: #45% chance to inherit a chromosome from parent 1
                child_chromosome.append(gp1)
            elif prob < 0.90: #45% chance to inherit a chromosome from parent 2
                child_chromosome.append(gp2)
            else: #10% chance to inherit a randomly mutated gene
                child_chromosome.append(self.mutated_genes())
        return Individual(child_chromosome)
    def cal_fitness(self): #returns the fitness score
        fitness = 0
        temp = self.chromosome
        p = list(temp)
        p.reverse()
        T_score = 0
        R_score = 0
        for Lambda in range(500,5000): #for all wavelength in [500,5000]
            M = boundary_matrix(n[int(p[0])],n[0])#using the TMM to calculate the optical property of the design
            M = np.matmul(M,propagation_matrix(phaseangle(int(p[0]),Lambda)))
            l = p[0]
            for item in p[1:]:
                if item != l:
                    M = np.matmul(M,boundary_matrix(n[int(item)],n[int(l)]))
                    M = np.matmul(M,propagation_matrix(phaseangle(int(item),Lambda)))
                elif item == l:
                    M = np.matmul(M,propagation_matrix(phaseangle(int(item),Lambda)))
                l = item
            M = np.matmul(M,boundary_matrix(n[0],n[int(p[-1])]))
            t = 1/M[1,1]
            if Lambda <= 1780:
                T_score += BB[Lambda-500]*np.absolute(t)**2 #Tansmission score (pass band)
            if Lambda > 1780:
                R_score += BB[Lambda-500]*(1-np.absolute(t)**2) #Reflection score (bandgap)
        fitness = 2/(T_score/idealT + R_score/idealR) # Compare the score to an ideal filter
        return fitness
def main():
    global POPUL_Size
    generation = 1
    found = False
    population = []
    for _ in range(POPUL_Size):
        gnome = Individual.create_gnome()
        population.append(Individual(gnome))
    while generation <= 200:
        population = sorted(population, key = lambda x:x.fitness)
        if population[0].fitness <= 0: #set your target fitness score here.We aim for lowest (numerically) fitness score so we set it as 0.
            found = True
            break
        new_generation = []
        s = int((10*POPUL_Size)/100)
        new_generation.extend(population[:s])
        s = int((90*POPUL_Size)/100)
        for _ in range(s):
            parent1 = random.choice(population[:50])
            parent2 = random.choice(population[:50])
            child = parent1.mate(parent2)
            new_generation.append(child)
        population = new_generation
        print("Gen: {}\tSolution: {}\tFitness Score: {}".
              format(generation,
              "".join(str(population[0].chromosome)),
              str(population[0].fitness)))
        generation += 1
    print("Gen: {}\tSolution: {}\tFitness Score: {}".
          format(generation,
          "".join(str(population[0].chromosome)),
          str(population[0].fitness)))
if __name__ == '__main__':
    main()
```

References

Aly, Arafa H., et al. "Biophotonic sensor for the detection of creatinine concentration in blood serum based on 1D photonic crystal." *RSC advances* 10.53 (2020): 31765-31772.

Barbieri, Enrico, et al. "Performance Evaluation of the Integration Between a Thermo–Photo–Voltaic Generator and an Organic Rankine Cycle." *Journal of engineering for gas turbines and power* 134.10 (2012).

Bauer, Thomas. *Thermophotovoltaics: basic principles and critical aspects of system design*. Springer Science & Business Media, 2011.

Becker, F. E., E. F. Doyle, and K. Shukla. "4th NREL Conf. on Thermophotovoltaic Generation of Electricity (AIP Conf. Proc. vol 460)." (1999).

Bett, Andreas W., et al. "III–V solar cells under monochromatic illumination." *2008 33rd IEEE Photovoltaic Specialists Conference*. IEEE, 2008.

Bradbury, Kyle, Lincoln Pratson, and Dalia Patiño-Echeverri. "Economic viability of energy storage systems based on price arbitrage potential in real-time US electricity markets." *Applied Energy* 114 (2014): 512-519.

Brown, Thomas G., et al. "The optics encyclopedia." *Physik Journal* 3.7 (2004): 22.

Celanovic, Ivan, et al. "Design and optimization of one-dimensional photonic crystals for thermophotovoltaic applications." *Optics letters* 29.8 (2004): 863-865.

Daneshvar, Hoofar, Rajiv Prinja, and Nazir P. Kherani. "Thermophotovoltaics: Fundamentals, challenges and prospects." *Applied Energy* 159 (2015): 560-575.

De Pascale, Andrea, et al. "Integration between a thermophotovoltaic generator and an Organic Rankine Cycle." *Applied Energy* 97 (2012): 695-703.

Dyachenko, Pavel N., et al. "Controlling thermal emission with refractory epsilon-near-zero metamaterials via topological transitions." *Nature communications* 7.1 (2016): 1-8.

Ferrari, C., et al. "Overview and status of thermophotovoltaic systems." *Energy Procedia* 45 (2014): 160-169.

Fraas, Lewis, et al. "Antireflection coated refractory metal matched emitters for use with GaSb thermophotovoltaic generators." *Conference Record of the Twenty-Eighth IEEE Photovoltaic Specialists Conference-2000 (Cat. No. 00CH37036)*. IEEE, 2000.

Fraas, L. M., et al. "Thermophotovoltaic system configurations and spectral control." *Semiconductor Science and Technology* 18.5 (2003): S165.

Fourspring, Patrick M., et al. "Thermophotovoltaic spectral control." *AIP Conference Proceedings*. Vol. 738. No. 1. American Institute of Physics, 2004.

Hernandez, Adrian Abdala Rendon. *Design, modeling and evaluation of a thermo-magnetically activated piezoelectric generator*. Diss. Université Grenoble Alpes, 2018.

Jena, S., et al. "Tunable mirrors and filters in 1D photonic crystals containing polymers." *Physica E: Low-dimensional Systems and Nanostructures* 114 (2019): 113627.

Khosroshahi, Farhad Kazemi, Hakan Ertürk, and M. Pınar Mengüç. "Optimization of spectrally selective Si/SiO2 based filters for thermophotovoltaic devices." *Journal of Quantitative Spectroscopy and Radiative Transfer* 197 (2017): 123-131.

Kumar, Manoj, et al. "Genetic algorithm: Review and application." *Available at SSRN 3529843* (2010).

Laroche, Marine, Rémi Carminati, and J-J. Greffet. "Near-field thermophotovoltaic energy conversion." *Journal of applied physics* 100.6 (2006): 063704.

Li, Zhi-Yuan, and Lan-Lan Lin. "Photonic band structures solved by a plane-wave-based transfer-matrix method." *Physical Review E* 67.4 (2003): 046607.

Liang, Peng, et al. "Boron implanted emitter for n-type silicon solar cell." *Chinese Physics B* 24.3 (2015): 038801.

Mao, Lei, and Hong Ye. "New development of one-dimensional Si/SiO2 photonic crystals filter for thermophotovoltaic applications." *Renewable Energy* 35.1 (2010): 249-256.

Menna, P., G. Di Francia, and V. La Ferrara. "Porous silicon in solar cells: a review and a description of its application as an AR coating." *Solar Energy Materials and Solar Cells* 37.1 (1995): 13-24.

Nunes, Pedro S., et al. "Refractive index sensor based on a 1D photonic crystal in a microfluidic channel." *Sensors* 10.3 (2010): 2348-2358.

Pirvaram, Atousa, et al. "Evaluation of a ZrO2/ZrO2-aerogel one-dimensional photonic crystal as an optical filter for thermophotovoltaic applications." *Thermal Science and Engineering Progress* 25 (2021): 100968.

Qiu, K., and A. C. S. Hayden. "Development of a novel cascading TPV and TE power generation system." *Applied Energy* 91.1 (2012): 304-308.

Stone, K. W., et al. "2nd NREL Conf. on Thermophotovoltaic Generation of Electricity (AIP Conf. Proc. vol 358)." (1996).

You, Chun-Yeol, and Sung-Chul Shin. "Generalized analytic formulae for magneto-optical Kerr effects." *Journal of applied physics* 84.1 (1998): 541-546.

Zhan, Tianrong, et al. "Transfer matrix method for optics in graphene layers." *Journal of Physics: Condensed Matter* 25.21 (2013): 215301.

Airworthiness Analysis of the Blended Wing Body Configuration by Using ANSYS Fluent as an Investigation Tool: SAX-40 as an Example

Yifan Wang

Author Background: *Yifan Wang grew up in China and currently attends the High School Affiliated to Shanghai Jiao Tong University in Shanghai, China. His Pioneer research concentration was in the field of engineering/mathematics and titled "Applied Mathematics for Engineers."*

Abstract

The blended wing body is a new kind of aircraft configuration which was conceptualized in 1988 by Rawdon, Lieback and Page to answer the question of whether there is an aerodynamic renaissance for long range transport. The blended wing body is a configuration that "stirred" the characteristic of a swept wing jetliner and fly wing body. Unlike conventional swept wing jetliners, not only the outer wing section can generate lift, but the fuselage of the blended wing body can also serve as a lift device; unlike a pure wing body, the blended wing body has a "fuselage" in its structure composition, but a pure wing body does not have a related structure. The outstanding aerodynamic performance made this configuration stand out among many other creative designs such as oblique wings and truss-braced wings. Furthermore, the blended wing body is a more environmentally friendly one compared to the conventional designs. The configuration also allows many of the latest technologies to be used, such as laminar flow technology, thrust vectoring and boundary layer ingestion. In this essay, ANSYS Fluent as a novel tool is used for the 2-D analysis of one of the fundamental airfoils of SAX-40 blended wing body aircraft. The pressure coefficient, lift coefficient, drag coefficient and lift to drag ratio as parameters are analyzed based on the computational fluid dynamics simulation carried out in ANSYS Fluent. A proposal is used for 3D analysis of whole airframe since the centerbody airfoil cannot be depicted. The following analysis of the center-body airfoil and 3-D analysis of airframe is based on existing data and graphs from other related essays. The study investigates how the blended wing body achieves longitudinal stability and aerodynamic efficiency, and aerodynamic performance characteristics of the two airfoils under different angles of attack are carried out as well.

1. Introduction

Since the Wright Brothers managed to make the first manned aircraft fly into the sky under control in 1903, various airplanes have enabled human beings to get rid of the control from gravity and cast their vision to the deep blue in the sky. Nowadays the majority of aircraft in service use the conventional "tube-and-wing" structure, and airplane manufacturers seem to be loyal to this design. The first flight of the Boeing B-47 Stratojet Bomber (Fig. 1) in December 1947 marked the appearance of modern tube-and-wing aircrafts. Used in both commercial and military fields, the technology of tube-and-wing aircrafts has been proven to be mature and reliable, but it cannot be the proof to its perfection: there are still aspects such as operation noise and fuel efficiency performance for further improvement. However, nowadays almost all enhancements on the aeronautical performance of the airplanes are attributed to the iteration of the engines or the propulsion system, indicating the reach of limit for the conventional aircraft configuration (Okonkwo & Smith, 2016).

Figure 1. *B-47 Stratojet Bomber. This airplane explores the design of swept wing with podded engines, and a fuselage is placed at the centerline of the aircraft. This design is also known as "wing and tube" configuration (Leone, 2019)*

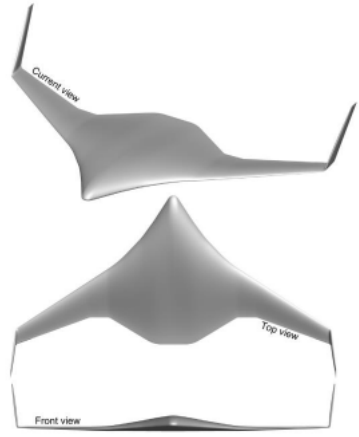


Figure 2. *A plane view of the blended wing body (Dehpanah et al., 2015)*

To answer the question raised by Dennis Bushnell about the potential renaissance of advanced aerodynamic configuration design (Larrimer, 2020), the Blended Wing Body (hereinafter referred to as BWB) configuration shown in the above Figure 2 was conceptualized in 1988. The configuration was further developed under the great contribution from three principal developers: Blain K. Rawdon, Robert H. Lieback, Mark A. Page (Larrimer, 2020). According to Timothy Risch, NASA Dryden Flight Research Center X-48 project manager,
"A blended wing configuration is characterized by an overall aircraft design that provides minimal distinction between wings and fuselage, and fuselage and tail. The blended wing configuration closely resembles a flying wing configuration but concentrates more volume in the center section of the aircraft than does the

traditional flying wing" (Larrimer, 2020).

In short, the BWB configuration can be seen as the fusion of conventional aircraft configuration and the flying wing configuration (Dehpanah & Nejat, 2015). Since the BWB does not have a horizontal tail for pitch control, its operation noise is largely reduced. The omission of a horizontal tail also contributes to the reduced total wetted area and leads to the reduction of total drag. Thus, the Lift-to-Drag ratio of the BWB is about 20% higher than that of conventional aircraft configuration, and the fuel burn per seat is 27% lower than conventional aircraft, which provides the BWB concept with longer range and larger payload capacity (Okonkwo & Smith, 2016; Dehpanah & Nejat, 2015; Jung Hoe & Nik Mohd, 2014; Qin et al., 2004). What is more, the airfoil-shaped fuselage for the BWB is also a lift device, so there is a spanwise lift distribution established by the BWB, which largely increases its aerodynamic efficiency and cruise performance.

However, there are still many problems that need to be solved when considering the BWB to be the next mainstream aeronautical configuration. In the conventional aircraft configuration, the function of each structure is independent, so the requirements for the multidisciplinary design were not very strong. The design for the conventional configuration can be modular, while the BWB is a highly integrated design: a change in one structure can influence a set of other structures, so it is necessary to use a multidisciplinary design optimization platform for future BWB development. One of the most well-known multidisciplinary design optimization methods is the Wing Multidisciplinary Optimization Design code (known as WingMOD) developed and used by Boeing. Nowadays, computational fluid dynamics is frequently used in the aircraft design with the computers' increasing calculation power, but the high fidelity still leads to high computational costs.

Another challenge is about the tailless nature of the BWB configuration: although the omission of a horizontal tail can solve the noise problem and enhance aerodynamic performance, it also causes concern in longitudinal pitch control method (geometry of aircraft principal axes is shown in Fig. 4). The omission of vertical tails, on the other hand, will affect the reliability of yaw stability and control (Wildschek, 2009). According to Wildschek et al., the conventional design of the trailing edge and winglet flaps is not effective for the BWB due to its height limitation, which will affect the total winglet areas (Wildschek, 2009). As a solution, he proposed "crocodile flaps" (Fig. 3) that separate the whole tailing edge flaps horizontally into two. This design allows the flaps to act as a conventional aileron when the upper and lower control surfaces are actuated in the same direction, and when external drag is required to provide yaw moment, the two control surfaces can move towards opposite directions. Wildschek has proven the feasibility of the crocodile flaps concept, but he also admits that much work needs to be done to test the capacity of structure under "real mission conditions": the structure is pushed towards its limit to achieve the expected aerodynamic function (Wildschek, 2009). The highly integrated nature of the blended wing body requires extreme structural complexity and strength. In order to take advantage of the aerodynamic efficiency the blended wing body, the shape of the airframe needs to be carefully designed.

In this essay, ANSYS Fluent as a novel Computational Fluid Dynamics simulation tool is used for analyzing the aerodynamic performance of the blended

wing body (in this essay SAX-40 is selected as the target airframe). Airfoils are selected with reference to the SAX-40 information for 2-D simulation. A proposal follows to analyze the center-body airfoil and 3-D model of the blended wing body.

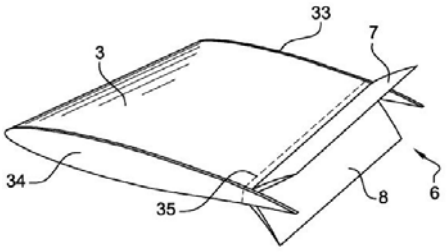


Figure 3. *One of the design graphs of crocodile flaps; note that there are two separate flaps at the trailing edge (Cazals et al., 2013)*

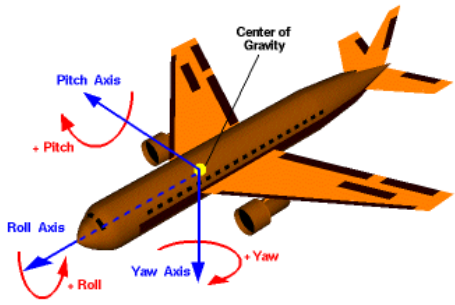


Figure 4. *Geometry of aircraft principle axes (Nonea et al., 2021)*

2. An overview of history on the Fly Wing and the Blended Wing Body

2.1. Before the Blended Wing Body

2.1.1 Before WWII: Germany as the Center of Development

When referring to the development of the blended wing body concept, the pure wing body configuration cannot be ignored: the close relationship between the two configurations is unveiled by their similar appearance. Human being's first concept of the wing body can be traced back to 1910 when Hugo Junker came up with the idea of a pure metal-built fly wing structure (Fig. 5) with passengers and the engine hidden within its outer skin. It is seen as the earliest concept of the flying wing (Larrimer, 2020).

The early blended wing body concept also relates to the engineers' pursuit of "Riesenflugzeuge" in the 20th century, basically giant planes. From Zeppelin-Staaken R.VI bomber that took the first flight in 1916 to Me 323 (Fig. 6) Gigant built in World War II and the biggest modern plane, Antonov An-225 Mriya (Fig. 7), they are all milestones that engineers took towards larger planes. Junker also made several attempts to build giant airplanes: Junker et al. developed the G40 trans-Atlantic seaplane project during in 1920s, and also pushed forward its land-based derivative, which is named G38 (Fig. 8). There were a total of 8 G38 airplanes being built, 2 prototypes constructed in Germany in 1929 and 6 by Mitsubishi. The biggest change Junker made about the G38 airplane is its capacity of carrying 6 people within its wings. The thought of making the wing a part of the fuselage is similar to the concept of blending the wing.

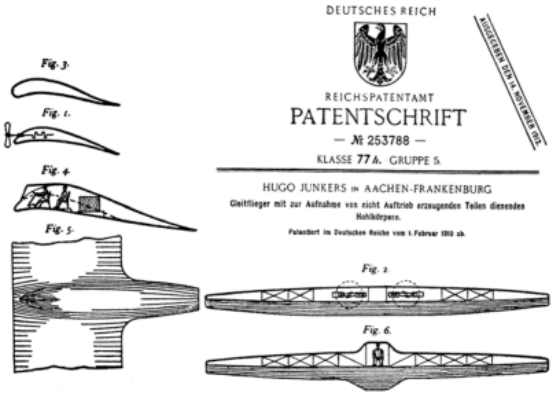


Figure 5. *The design of Junker's pure wing body airplane. All the passengers as well as the engine are placed under the metal skin within the airplane (Okonkwo et al., 2016)*

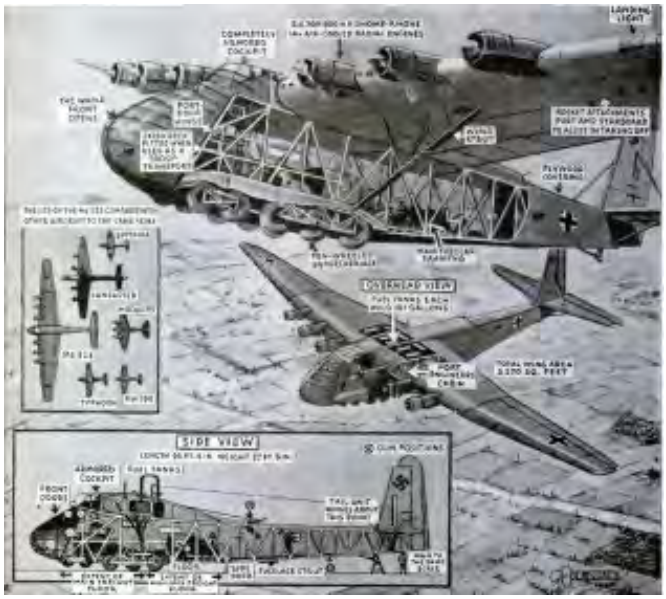


Figure 6. *The internal composition graph of Me323 and size comparison with other existing fighters and bombers in World War II (2022)*

Figure 7. *Antonov 225 Mriya. It was designed for carrying Buran-class orbiters. Unfortunately, the only A-225 was destroyed in the Battle of Antonov Airport (Lesiv et al., 2022).*

Figure 8. *G38 trans-Atlantic 4-engine landplane (Cluett, 2022)*

At the end of World War II, airplane technology and design ideas were decades ahead of their time, especially for jet-engine-powered air-fighters. With the groundbreaking design of using jet engines and swept wings (Christopher, 2013), the Messerschmitt Me 262 became the first operational jet-powered air-fighter in human history. Along with the swept wing, German engineers also tried to combine other configurations with jet engines, and thus the Horten Ho 229 (Fig. 9) bomber was developed by Walter Horten and Reimar Horten. It was a bomber designed to satisfy the "3×1000" requirement (carry 1,000 kilograms of bombs travel a distance of 1,000 kilometers with a speed of 1,000 kilometers per hour; Boyne, 1994). If the Horten brothers used the popular flight configuration at that time, the requirement could not be fulfilled even with the help of the latest Junkers Jumo 004B turbojets. The easiest one of the three requirements, the speed requirement, still could not be reached without extra fuel consumption. The final result was the adoption of the fly-wing scheme due to its low drag and high cruise efficiency characteristics. This design yielded the first aircraft that combines the jet engine and flying wing together. There were a total of 6 iterations of Ho 299 bombers designed during World War II, but due to the gradual defeat of Germany in Europe, the Ho 229 was not mass produced. When World War II was about to come to an end, Operation Paperclip initiated by US military brought the prototype of Ho 229 to native America, and the only prototype of the Ho 220 left now is in the Smithsonian National Air and Space Museum's Paul E. Garber Restoration Facility in Suitland, Maryland, U.S.

2.1.2. The United States Led the Trend

With the end of World War II, the forefront of fly wing research shifted from Germany to America. In the 1950s, several fly wing projects were carried out by the USA. However, hardly any of them survived: "in every case authorities rejected these in favor of more conventional wing-body-tail designs" (Larrimer, 2020). Among those projects, the best known is the four-engine-piston-powered XB-35 (Fig. 10) family and its 8-engine-jet-powered derivatives developed by Northrop Cooperation. However, during the test flights, a severe longitudinal stability problem was exposed: the tailless nature of the fly wing configuration requires the pilot to actively intervene in aircraft control when entering the compressible flow. This problem resulted in the cancellation of the XB-35 project.

The cancellation of XB-35 made the US Air Force turn to concepts of passengers, freight and cargo located in the wings instead of in the conventional fuselage. The longitudinal pitch problem was not solved until the release of the Northrop B-2A Spirit bomber. The digital electronic flying control technology equipped by the B-2A proved the feasibility of computer controlled flight (Larrimer, 2020), exploring the possibility of a fly wing configuration.

Figure 9. *Center-body of Ho 229 Komet Fighter with its removed wings placed on the right in the Smithsonian National Air and Space Museum's Paul E. Gerber Restoration Facility in Suitland, Maryland, U.S (Dowling, 2017)*

Figure 10. *XB-35 fly wing aircraft (Par et al., 2014)*

2.2. Conceptualization and Development of the Blended Wing Body

2.2.1 Answer to Dennis Bushnell's Problem

In 1988, Dennis Bushnell from NASA asked the following question:
“Is there an aerodynamic renaissance for the long-haul transport?” (Larrimer, 2020)

Bushnell mainly focused on the innovation in aerodynamic design, and engineers gave many creative new ideas: the blended wing body presented by McDonnell Douglas and now Boeing’s Robert H. Lieback (Fig. 11), the oblique wing presented by R.T. Jones from NASA, and Ames’ legendary truss-braced wings presented by Werner Pfenninger from NASA Langley, etc. (Larrimer, 2020). In 1994, NASA initiated the Advanced Concepts for Aeronautics Program, and a 3-million-dollar, 3-year-long contract known as NASA-20275 was given to McDonnell Douglas, leading to a detailed comparison among three different configurations: a swept wing jetliner, a pure fly wing and a blended wing body (Larrimer, 2020). In the experiment, the blended wing body established a high cruise lift-to-drag ratio up to 27.2 and high fuel burn efficiency. According to the engineers, “(the blended wing body) indicates that the blended-wing-body configuration is the superior performer.” In 1997, McDonnell Douglas merged Boeing, but at that time Boeing was unsure about whether the blended wing project was worth their further investment. It was NASA that convinced Boeing of the potential and feasibility of the blended wing body project.

In 1998, three principle concept formers of the blended wing body made further definition and added aerodynamic design to the configuration. Boeing did

a series of experiments with models at different scales, e.g. BWB-6, BWB-17, X-48B, etc., (Okonkwo & Smith, 2016; Larrimer, 2020), and thus paved the way for the 450-passenger full-scale BWB-450 project design (Fig. 12). BWB-450 is a commercial aircraft designed for a payload of 468 passengers separated in 3 classes and a designed range of up to 7750 nautical miles. The BWB-450 is developed under the WingMOD multidisciplinary design optimization platform, and the platform allows more than 20 design conditions to be considered during the design process (Wakayama, 2000). There are many novel technologies applied to this aircraft, including laminar flow technology, jet flaps, distributed propulsion, boundary-layer ingestion, etc. (Okonkwo & Smith, 2016; Larrimer, 2020; Dehpanah & Nejat, 2015; Qin et al., 2004; Liebeck, 2002).

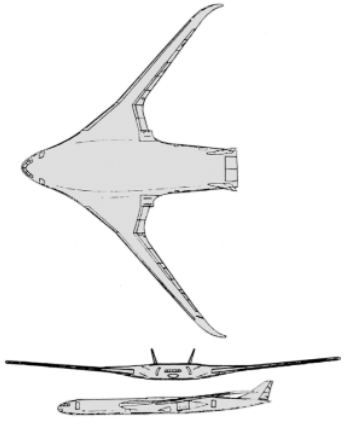


Figure 11. *First generation design of the blended wing body developed by Lieback*

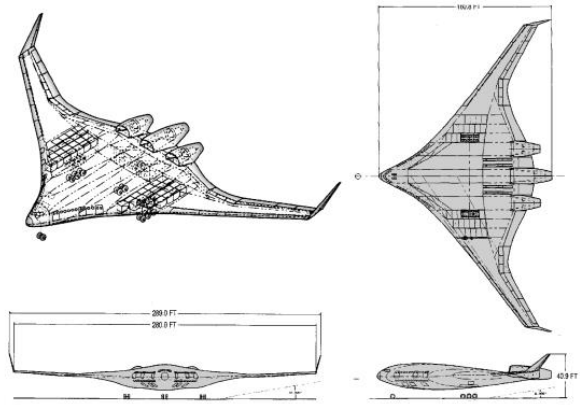


Figure 12. *BWB-450 baseline geometry (Liebeck, 2016)*

2.2.2 Various Blended Wing Body Projects

In the 21st century, increasing attention has been paid to the wing body configuration family, and as a result, an increased number of colleges, companies and countries have initiated their own blended wing body programs. European countries and Russia have all had similar projects and designs: Very Efficient Large Aircraft project (VELA), New Aircraft Concept Research (NACRE) from the European Union (EU) framework; Silent Aircraft eXperimental passenger aircraft (SAX) series as a product of the combined work of Cambridge University and Massachusetts Institute of Technology (MIT); The Zhukovsky Central AeroHydrodynamic Institute (TsAGI) program initiated by Russia (Okonkwo & Smith, 2016).

2.2.2.1 VELA Project

The Very Efficient Large Aircraft (VELA) project is a 3-year-long project that yields two distinct airframes (Okonkwo & Smith, 2016). The two aircrafts, named VELA 1 and VELA 2, share the same design goals, and many similarities can be found between the two airframes. The most obvious one is the outer wing section of the two baseline airframes: all the wings in the outer wing section are highly

similar to the conventional swept wing ones, and both airframes have 4 podded engines as the power source. In addition, both airframes use two vertical tails for yaw control. The two baseline designs only vary in the placement and the blending of the outer wing section. Based on the two baseline airframes, VELA 3 (Fig. 13) was conceived with the blending part of the outer wing placed in the middle of the central fuselage. The VELA3 is designed to carry 750 passengers and travel 7500nm at cruise speed of 0.85 Mach (Okonkwo & Smith, 2016).

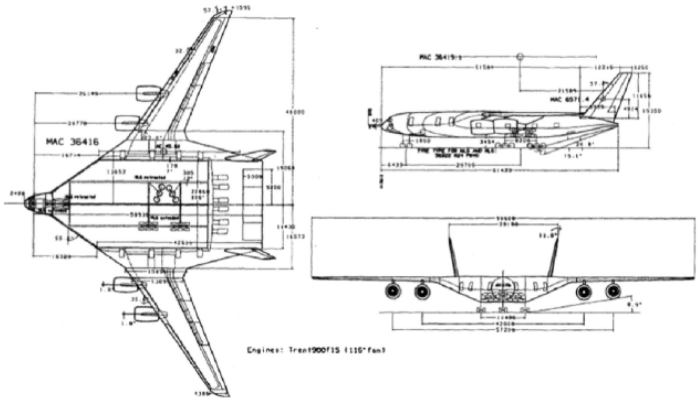


Figure 13. *The finial output of VELA project named VELA 3 (Okonkwo & Smith, 2016)*

2.2.2.2 NACRE Project

The New Aircraft Concept Research (NACRE, shown in Fig. 14) is a 4-year program started in 2005 led by Airbus, and 13 countries were involved in the design process (Okonkwo & Smith, 2016). NACRE further developed the blended wing body concept through its Passenger-driven Fly Wing (PFW) baseline airframe (Okonkwo & Smith, 2016). There are two generations of PFW baseline airframe produced by this program. The PFW airframe is based on the VELA-3 airframe, and PFW-1 enhances the aerodynamic performance of VLEA-3 by applying a wing twist to VELA-3; PFW-2 further changes the location of engines and the blending design of PFW-1 (Okonkwo & Smith, 2016; Okonkwo, 2016).

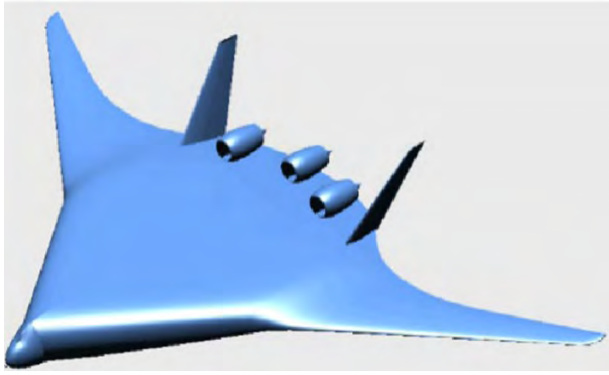


Figure 14. *NACRE PFW-2, the second-generation aircraft of NACRE program (Okonkwo & Smith, 2016)*

2.2.2.3 SAX Project

The Silent Aircraft eXpenrimental passenger aircraft (SAX, shown in Fig. 15) is an aircraft jointly developed by the Massachusetts Institution of Technology and Cambridge University (Hileman et al., 2007). There are many versions of the airplane yielded in the project, and the latest version is named SAX-40. Many novel technologies of noise control and aerodynamic design were applied to the project, giving the aircraft a far-field noise reduction for over 20bB. What is more, since the noise reduction design will also help improve the aerodynamic performance and thus reduce fuel consumption, fuel efficiency is increased at the same time. Unlike other projects, SAX-40 is planned to enter service in 2030, but there are still many problems waiting to be solved: the manufacture of the unique shape of the airframe and vibration caused by the non-uniform inlet flow in boundary layer ingestion (Okonkwo & Smith, 2016; Hileman et al., 2010).

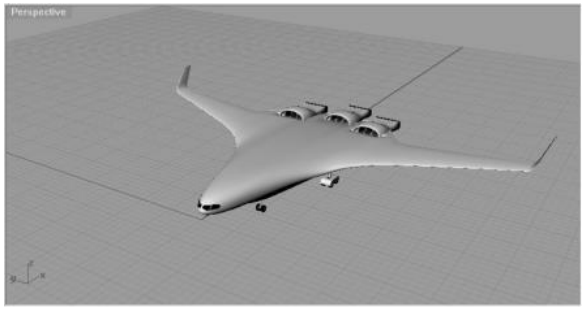


Figure15. *SAX-40 airframe concept (Hileman et al., 2010)*

2.2.2.4 TsAGI Project

The Zhukovsky Central AeroHydrodynamic Institute (TsAGI, shown in Fig. 16) project used the results from the VELA project and conducted research including 4 airframes to initiate the critical concept in the blended wing body design (Okonkwo & Smith, 2016). The results turned out to be the integrated wing body, lift body configuration and pure flying wing. A series of experiments were planned to find the optical design among the three concepts, and the integrated wing body (Fig. 16) turned out to be the one.

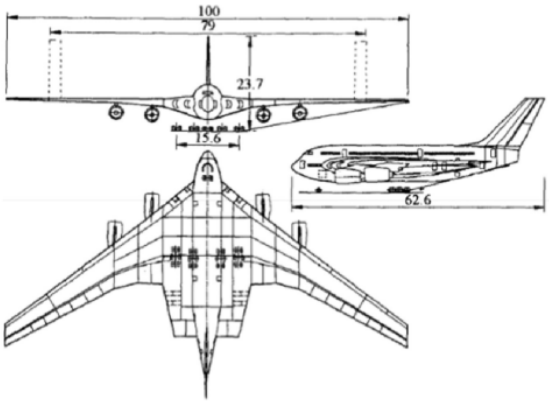


Figure 16. *TsAGI Integrated Wing Body design (Okonkwo & Smith, 2016)*

In this section, a numerical analysis method based on airfoils similar to the SAX-40 is carried out in the Computational Fluid Dynamics tool ANSYS. The methodology introduction includes the reliability evaluation of ANSYS and ANSYS-Fluent program, introduction to the library used in baseline airfoil selection and the convergence analysis as well (Rahimi et al., 2014). The introduction towards the viscous model used for analyzing will also be included.

3.1 ANSYS

ANSYS, founded in 1970, is a whole product-life-cycle-covering engineering simulation software. By using Finite Element Analysis that can predict the real-life state of a structure, ANSYS provides simulation models covering structure, electronics, temperature distribution, fluid dynamics, etc. Most functions of ANSYS rely on the ANSYS workbench, a platform that integrates all the modules. The workbench allows users to construct a complex assembly from individual parts and analyze them using different modules depending on the demand. In this essay, aerodynamic analysis will be carried out, so the ANSYS Fluent program module will be used. According to the official website, the "fast pre-processing and faster solve time" expands the capacity and possibility of CFD software while ensure its quality and professionalism at the same time. There is a great number of existing essays that have utilized this analysis software for aerodynamic analysis: Meyers et al. used the ANSYS-Fluent program for the hydrodynamic analysis of the modified trailing edge of an underwater glider wing (Meyers & Msomi, 2021); González et al. used the ANSYS-Fluent program to analyze the effect of incorporating protuberances in the leading edge to the aerofoil NACA-0012 (González & Hinojosa, 2019); Murariu et al. used the ANSYS-Fluent program for fluent simulation for the fly-wing unmanned aerial vehicle UAV prototype; Cayiroglu et al. used the ANSYS-Fluent program for the wing simulation of a private jet plane (Cayiroglu & Kilic, 2017).

3.2 Geometry Definition

As mentioned before, there are multiple blended-wing body projects carried on in different countries by various institutions, and none of them can grant the writer access. Therefore, the writer chooses the most open one, SAX-40, as the baseline geometry. According to Jones et al., the SAX-40 3-D airframe model is created with reference to 4 discrete airfoils as shown in Figure 17, and airfoils with fraction numbering are the intermediate ones depicted by using linear interpolation to merge the adjacent airfoils (Jones, 2006; Zhonghua et al., 2021; Hileman et al., 2010). The thickness to chord ratio and wing twist along the wing is shown in Fig. 18. The whole airframe is divided into four distinct airfoils, in response to center-body, inner wing, outer wing and winglet (Jones, 2006; Zhonghua et al., 2021). However, according to Sargeant et al. and Hileman et al., who are respectively from Cambridge University and the Massachusetts Institute of Technology, the two schools that initiated the SAX project, there are only two fundamental airfoils used in SAX design: the center-body one and the outer wing one, which are airfoils 1 and 3. So, the aerodynamic analysis in this paper will be mainly based on these two airfoils.

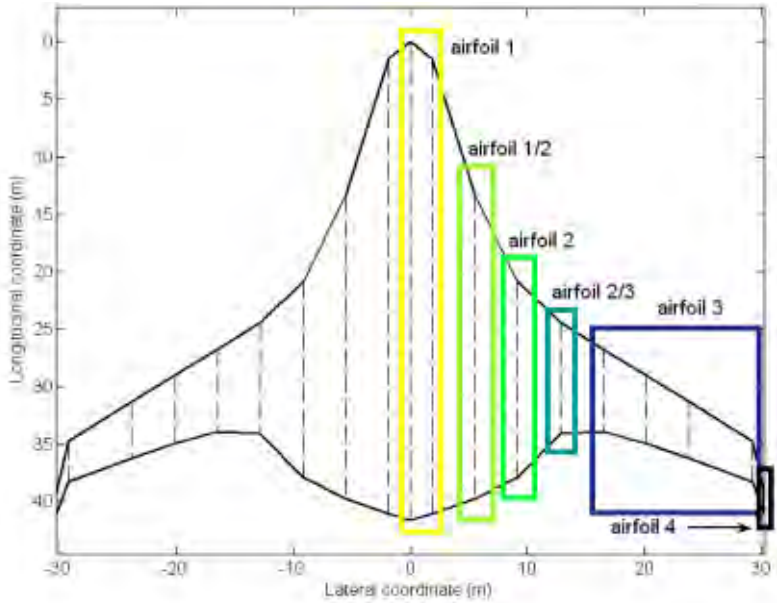


Figure 17. *Airfoil spanwise location (Jones, 2006)*

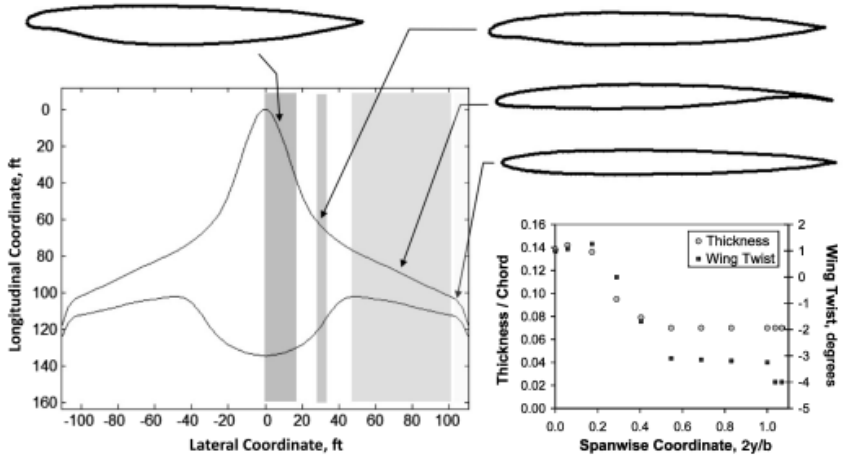


Figure 18. *SAX-40 airfoil section and thickness, wing twist spanwise distribution (Hileman,2010)*

Unlike conventional airfoils, all the airfoils used for SAX-40 are specially designed: each airfoil is divided into 5 segments and defined by Bézier Spline (Jones, 2006), so they cannot be found within the existing airfoil code. The writer is not able to precisely plot the airfoil with the given information, thus it is not applicable to use the precise coordinates of the origin airfoil for modelling and aerodynamic evaluation. If the writer could access the detailed airfoil coordinates of SAX-40, experiments with higher fidelity could be carried out. For the current situation, the airfoil under analysis will use similar baseline airfoils in the existing

airfoil library provided by University of Illinois Urbana-Champaign (UIUC). The airfoil 1 in the center-body and inner wing is the one with a high thickness to chord (basically the length of the wing/airfoil) ratio and leading-edge carving which shifts the aerodynamic center forward (Okonkwo & Smith, 2016; Jones, 2006; Zhonghua et al., 2021; Hileman et al., 2010; Sargeant et al., 2010). Since there is no existing airfoil in the UIUC airfoil library, the two airfoils will use the existing data and analysis for aerodynamic evaluation. Airfoil 3 in Figure 14 is a typical modern supercritical airfoil with a thickness to chord ratio of 7% (Zhonghua et al., 2021; Hileman et al., 2010). In the 1960s, NASA started the preliminary study on modern supercritical airfoils, and thus developed NASA SC supercritical airfoil series. There are three generations of supercritical airfoil, distinguished through the number in the parentheses in the airfoil code. In this paper, the second-generation airfoil is chosen, and according to the thickness to chord ratio, NASA SC (2)-0406 airfoil (Figure 19) is chosen for analysis. The "04" means that the designed lift coefficient of the airfoil is 0.4 while the "06" represents maximum thickness to chord ratio of 6%. Geometries of NASA SC (2)-0406 are presented in the table 1 as following.

Table 1. *Geometry summary of NASA SC(2)-0406 airfoil*

	"NASA SC(2)-0406"
"chord length(mm)"	100
"maximum thickness(mm)"	6
"location of maximum thickness(%chord)"	35
"maximum camber(mm)"	0.6
"location of maximum camber(%chord)"	79

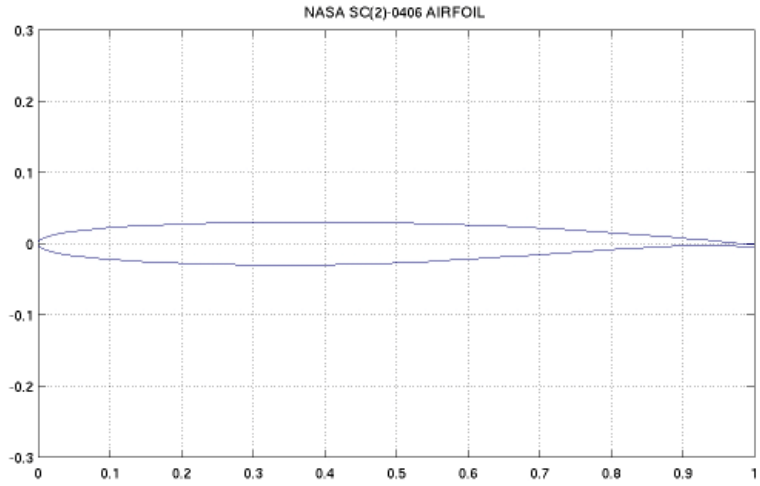


Figure 19. *2-D plot of airfoil based on NASA SC(2)-0406 airfoil (Selig, 1996)*

3.3 Viscous Model

The paper uses the Realizable k-epsilon (k-ε) Turbulence Model in the ANSYS-Fluent program for aerial simulation of the airfoils, and it is adapted from the Navier-Stokes equations and continuity equations shown as follows,

$$\rho\frac{Du}{Dt} = \rho\left(\frac{\partial u}{\partial t} + u\cdot\nabla u\right) = -\nabla p + \nabla\cdot\left\{\mu\left[\nabla u + (\nabla u)^T - \frac{2}{3}(\nabla\cdot u)I\right]\right\} + \rho g$$

Navier- Stokes equation (Kundu et al., 2016)

$$\frac{\partial}{\partial t}(\rho u) + \nabla\cdot(\rho u \otimes \mathrm{u}) = -\nabla p + \nabla\cdot\tau + \rho g$$

Continuity equation (Kundu et al., 2016)

The realizable k-ε turbulence model is a two-equation model that describes the airflow in turbulence conditions by using partial differential equations. The two differential equations are given below:

$$\begin{cases} \frac{\partial}{\partial t}(\rho k) + \frac{\partial}{\partial x_j}(\rho k u_j) = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_t}{\sigma_k}\right)\frac{\partial k}{\partial x_j}\right] + P_k + P_b - \rho\epsilon - Y_M + S_k \\ \frac{\partial}{\partial t}(\rho\epsilon) + \frac{\partial}{\partial x_j}(\rho\epsilon u_j) = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_t}{\sigma_k}\right)\frac{\partial \varepsilon}{\partial x_j}\right] + \rho C_1 S\epsilon - \rho C_2\frac{\epsilon^2}{k+\sqrt{v\epsilon}} + C_{1\epsilon}\frac{\epsilon}{k}C_{3\epsilon}P_b + S_\epsilon \end{cases}$$

where $C_1 = \max\left[0.43, \frac{\eta}{\eta+5}\right], \eta = \mathrm{S}\frac{k}{\epsilon}, \mathrm{S} = \sqrt{2S_{ij}S_{ij}}$

The realizable k-ε model is superior when the calculation of flow involves the situation of rotation, separation and recirculation.

3.4 Convergence Analysis and Boundary conditions

In order to ensure the effectiveness of meshing method, a convergence analysis is carried out based on the NASA SC (2)-0406 airfoil. The domain used in convergence analysis as well as the following investigation is composed of a semi-circle in the front at the leading edge of the airfoil following a rectangle (Figure 20). The radius of the semi-circle is 2 times the chord length of the airfoil, which is 200mm, and the length of the rectangle is 300mm, which 3 times the length of the chord length of the airfoil. The inlet air speed is set to 15 m/s, and the fluid in the CFD is chosen as air. The meshing method established in this paper mainly uses three built-in functions: meshing method, edge sizing and refinement. All the meshing methods are set to triangle ones, and edge sizing along with refinement varies as the independent parameter. The dependent parameter used is two figures: lift coefficient (Cl) and lift force (Fl). There are a total of five different meshing methods, and the one with the number of divisions up to 400 for edge sizing and a refinement number set to 3 produces the biggest number of elements as well as has the highest computational cost, so it is set as the base number. According to Table 2, all the errors of the first four results are below 10%, which is negligible for the low fidelity aerodynamic analysis, and the refinement number of 3 with number of divisions of 200 is chosen in consideration of

accuracy, stability as well as the computational cost.

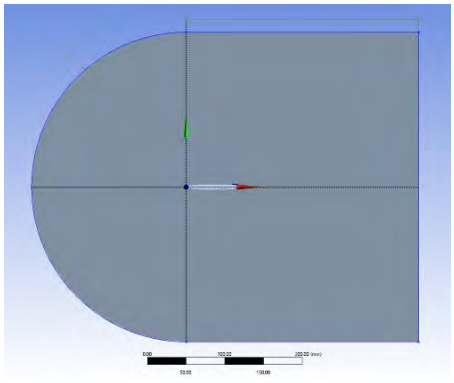


Figure 20. *The numerical domain used in the convergence analysis for the meshing method, R1 is two times of the chord length and H2 is three times of the chord length.*

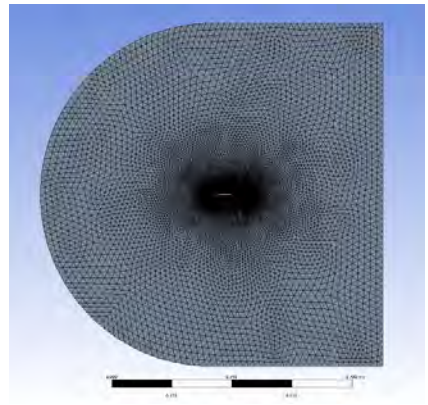


Figure 21. *The meshing produced with a number of divisions of 200 and a refinement number of 2*

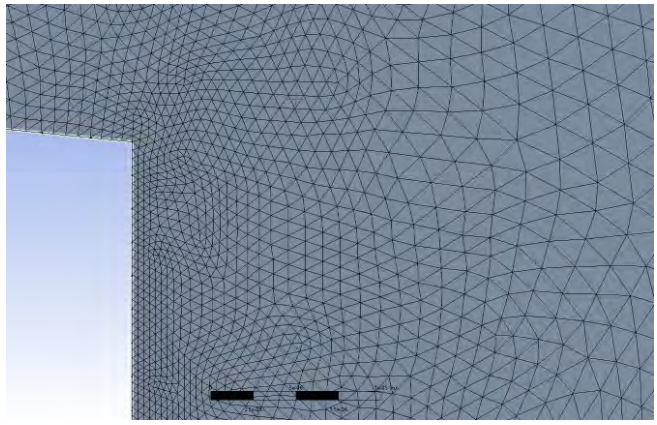


Figure 22. *A closer look at the meshing method: density of grid increases approaching the target airfoil.*

Table 2. *The result summary of the convergence analysis*

	Number of division	Refinement number	Number of elements	C_1	Error of C_1(%)	F_1	Error of F_1(%)
1	400	3	212418	0.044	0.	6.03	0.
2	200	3	133457	0.041	6.8	5.69	5.6
3	200	2	71612	0.04	10.	5.54	8.1
4	200	1	32077	0.041	6.8	5.64	6.4
5	200	0	8356	0.028	36.4	3.79	37.1

4. Results and Analysis

Hileman et al. give out the SAX-40 geometric parameters and the aerodynamic performance parameters of the SAX-40: the design cruise speed of SAX-40 is 0.8 Mach (equal to 274.4 m/s), outer wing twist is -3.25° (Fig. 18) and cruise Angle of Attack (the angle between the airfoil and direction of motion) is 2.7° (Hileman et al., 2010). The following analysis is done based on the parameters provided above, and the focus will be on stability characteristics as well as pressure distribution.

4.1 Pressure Coefficient Analysis on Outer-wing Airfoil

The NASA SC (2)-0406 airfoil is placed at different Angles of Attack varying from -2° to 8° with an interval of 2 degrees in ANSYS to give a full picture of the relationship between the Angle of Attack and the pressure coefficient distribution. According to the data given above, at cruise conditions, the Angle of Attack for the outer-wing airfoil is at -0.55°, so this specific Angle of Attack will be calculated and evaluated as well.

4.1.1 Airfoil at Cruise Condition

Fig.23 shows the pressure coefficient distribution along the chord length. From the graph, we can observe that nearly all the lift force is generated in the second half of the chord length, and the pressure above and under the airfoil is larger than the free-stream static pressure. The difference in pressure coefficient in the first half of the airfoil is nearly negligible, indicating no lift generated. At cruise conditions, the pressure center is near the trailing edge of the airfoil and creates a nose-down pitch moment, which fits the description of other related research (Hileman et al., 2010; Sargeant et al., 2010; Aprovitola et al., 2022).

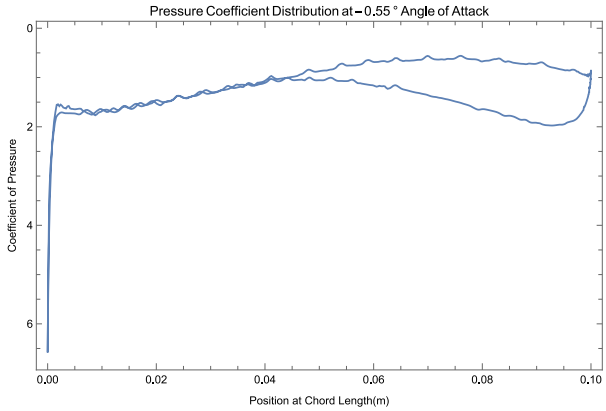


Figure 23. *Pressure coefficient distribution of the NASA SC (2)-0406 airfoil*

4.1.2 Airfoil at Different Angles of Attack

Fig.24 summarizes a total six graphs of pressure coefficient distribution along the chord length at Angles of Attack varying from -2° to 8°. From the set of graphs,

the trend of pressure coefficient distribution can be observed and analyzed. When the Angle of Attack is smaller than 4°, pressure at both upper and lower surfaces of the airfoil is larger than the free-stream static pressure, shown as a positive figure in the pressure coefficient. With the increase of the Angle of Attack, the pressure surrounding the airfoil begins to fall, and some parts of the pressure are even lower than free-stream static pressure when the Angle of Attack is larger than 4°.

What's more, with the variation of Angle of Attack, there are also changes in the lift distribution. At the first half of chord length, the pressure coefficient difference between the upper and lower surface of the airfoil increases with the increase of Angle of Attack. As lift is generated by the pressure difference between the upper and lower surfaces of the airfoil, lift generated at the first half of chord length gets bigger, indicating a positive relationship with the Angle of Attack. Compared to the variation at the leading edge of airfoil, change in lift at the trailing edge is not obvious. Thus, the center of pressure will shift forward as the Angle of Attack increases.

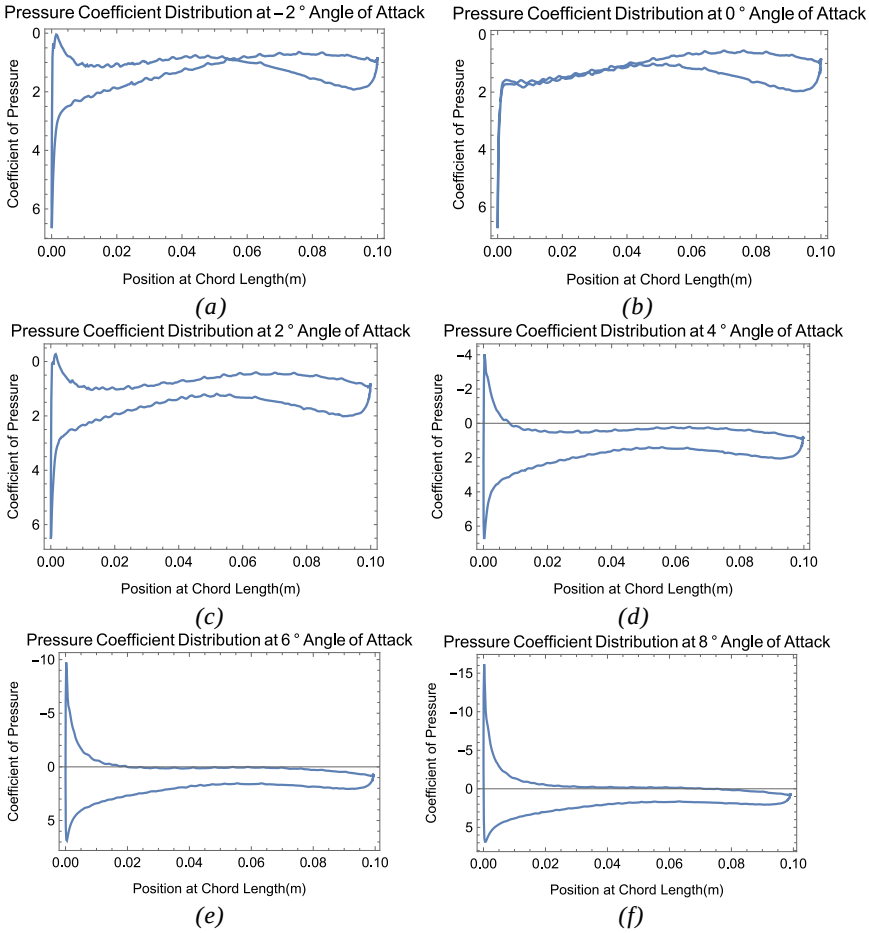


Figure 24. *Pressure coefficient distribution of the airfoil at different Angle of Attack at the cruise speed of 0.8 Mach*

4.2 Lift and Drag Coefficient Analysis

Fig. 25 depicts the lift coefficient and drag coefficient of the NASA SC(2)-0406 baseline airfoil with reference to the variation in the Angle of Attack. There is a general increase trend of lift coefficient and drag coefficient, but the slope of the two lines varies a lot: the gradient of lift coefficient is much larger than that of drag coefficient. Current data is not enough for the analysis of composition of drag and the reason behind those trends, but the aerodynamic efficiency can be discussed. Aerodynamic efficiency is indicated by the lift to drag ratio, which is calculated by the quotient of lift coefficient and drag coefficient (Cl/Cd). As Fig. 26 shows, lift to drag ratio is negative at -2° and increases to a peak of 9.61 at about 4° Angle of Attack, then generally decreases as the Angle of Attack further increases. According to the analysis above, at -0.55° cruise Angle of Attack, the aerodynamic efficiency of the airfoil does not reach a peak. However, related optimization to the baseline airfoil based on the aerodynamic efficiency can be carried out to further increase the performance of the airfoil (Qin et al., 2004; LI et al., 2012; Jones, 2006).

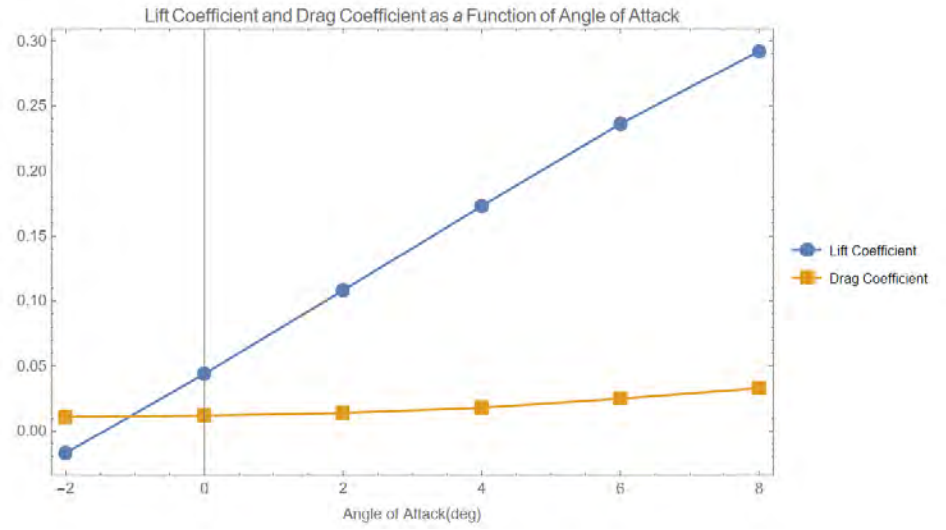


Figure 25. *Lift coefficient and drag coefficient as a function of Angle of Attack at the cruise speed of 0.8 Mach. The blue curve is the lift coefficient, and the yellow curve is the drag coefficient.*

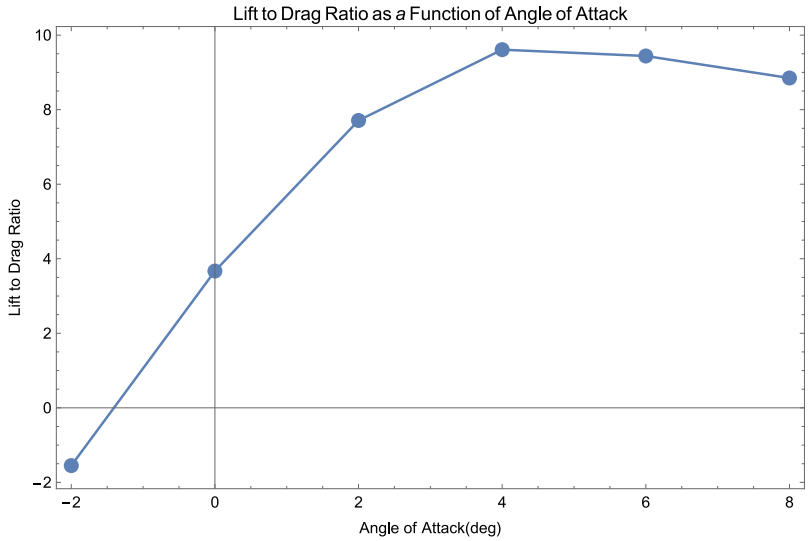


Figure 26. *Lift to Drag ratio as a function of Angle of Attack at the cruise speed of 0.8 Mach. The Lift to Drag ratio first increases till the Angle of Attack reaches 4° and then decrease.*

5. Discussion

5.1 Limitations of current analysis

In this paper, only the low-fidelity model is used for the simulation. Although it effectively lowers the threshold and computational cost, the data can only be used for preliminary analysis. Profound research requires higher fidelity models and a more advanced computer or computer network. Limited by the writer's knowledge background, calculation of more advanced parameters is not practical at the current stage. If the writer learns partial differential equations in college, detailed analysis of different kinds of drag and lift force can be added.

However, no matter how perfect the meshing method is or how advanced models are used, a 2-D analysis of the airfoil is not enough for the aerodynamic evaluation of the blended wing body configuration due to its lack of capability in having an overall picture of the airframe. Since the blended wing body is a highly integrated airframe, a specific airfoil or aerial design may bring a knock-on effect to the whole airframe. According to Hynes et al., the great reduction of $dc_l/d\alpha$ illustrates the incapability of 2-D methods in characterizing the pitch stability of whole airframe, so a 3-D method is required (Sargeant et al., 2010).

5.2 Future Research

5.2.1 Proposal: Center-body wing airfoil

Since the writer cannot access detailed depictions of the center-body airfoil, research data from the existing papers will be used to analyze the aerodynamic performance of the center-body airfoil. Fig. 27 shows the coefficient of pressure distribution of the whole airframe and the specific airfoils selected. The airframe used is the SAX-29, which is the former generation of the SAX-40. According to

Hileman et al., SAX-40 mainly optimized the outer-wing platform and wing twist based on low approach noise and high cruise fuel efficiency (Hileman et al., 2010), so when analyzing the center-body airfoil aerodynamic performance, it is similar to use the SAX-29 compared to SAX-40. From the pressure coefficient graph presented below, there is negligible pressure difference between the upper surface and lower surface near the trailing edge of the airfoil, indicating no lift generated. All the lift is generated in the front half of the chord length because of the presence of camber near the leading edge, and the second half is nearly unloaded. The center of pressure, as a result, is located at the front of the airfoil at about 35% chord length. This creates a nose-up pitching moment, which balances the nose-down pitching moment created by the outer-wing supercritical airfoils. Fig. 28 depicts the pressure distribution of the airfoil locating at 7.1% semi-span with different Angles of Attack: every 0.5° from 3° to 5° (Sargeant et al., 2010). The 7.1% semi-span is involved in the center-body section of the whole airframe and thus can act as a representation of the center-body airframes (Sargeant et al., 2010). There is not an obvious change in the pressure distribution of the airfoil, providing a possibility for the longitudinal pitch stability.

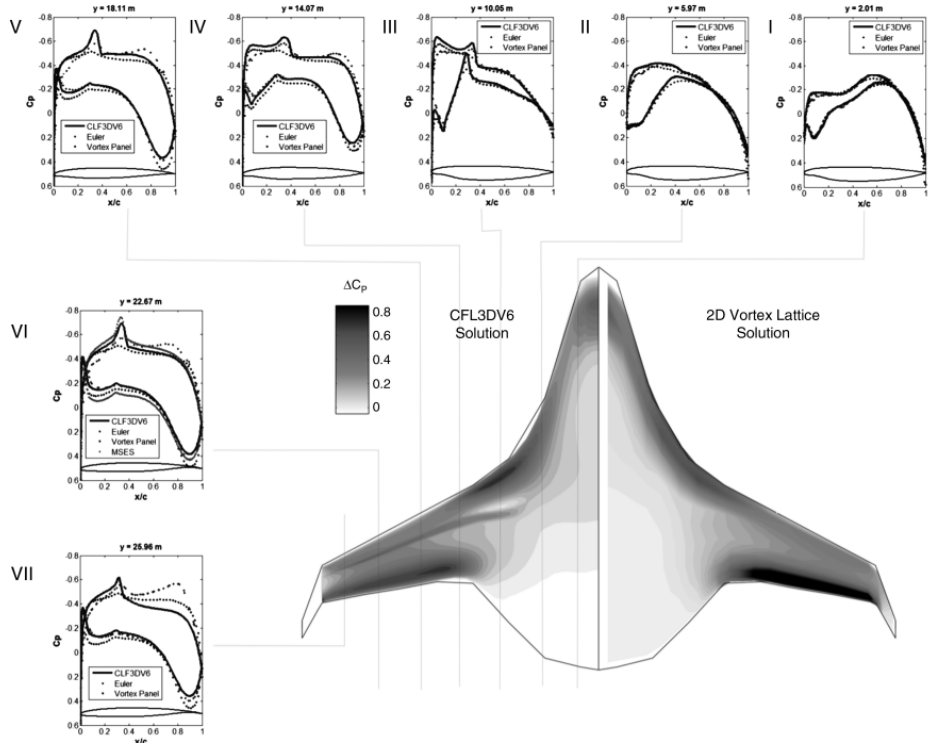


Figure 27. *Pressure distribution of the whole airframe and airfoils from different semi-span: each smaller graph represents the Pressure Coefficient graph at the semi-span indicated, and on each graph, there are three to four different marks indicating results generated by different models (Hileman et al., 2010)*

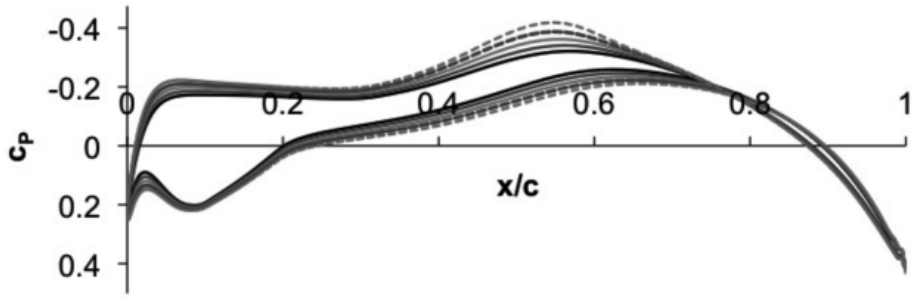


Figure 28. *Pressure coefficient of the airfoil at 7.1% semi-span at various Angle of Attack of SAX-29 calculated by using 3-D Navier-Stokes equation (Sargeant et al., 2010)*

5.2.2 Proposal: 3-D analysis of the blended wing body configuration

A 3-D model can be built by discretizing the whole airframe into several sections defined by specific airfoils and blending the adjacent sections with reference to wing twist (Jones, 2006). If the writer is able to build a 3-D airframe on the simulation software or gain direct access to the 3-D model of SAX-40, one could use the data from direct simulation for analysis, but at the present stage, data and graphs from other essays will be used for 3-D Reynolds averaged Navier-Stokes equations of the blended wing body configuration.

What is special about the blended wing body configuration is its spanwise lift distribution. Unlike a conventional tube-and-wing airframe, the airfoil-like fuselage also exerts lift-generation capability and acts as a lift generator. As Fig.29 illustrates, the lift coefficient is not high at the center due to the high thickness to chord ratio of center-body airfoil, but the longer chord length and the larger reference area makes the lift force generated by the center-body higher than that of the outer wing section. The lift distribution thus forms the elliptic spanwise lift distribution like Fig. 30 shows. According to the Equation:

$$\left(\frac{L}{D}\right)_{max} = \sqrt{\frac{\pi AR}{4kC_{d0}}} = b\sqrt{\frac{\pi}{kS_{d0}}}$$

($\boldsymbol{AR}$ represents aspect ratio, $\boldsymbol{k}$ represents induced drag factor, $\boldsymbol{C_{d0}}$ represents zero lift drag coefficient, $\boldsymbol{S_{d0}}$ represents reference area at 0 lift drag)

the elliptic spanwise lift distribution achieves maximum lift to drag ratio since the distribution has the least induced drag in subsonic cruise conditions (Okonkwo & Smith, 2016). However, in transonic conditions, the increased importance of wave drag prevents this elliptical from achieving optimal aerodynamic performance (Okonkwo & Smith, 2016). According to the research of Qin et al. who studied different spanwise lift distributions of outer wing sections in transonic cruise speeds: triangular and elliptic-triangular spanwise lift distribution is proposed (Okonkwo & Smith, 2016; Qin, 2002) (Fig. 31). The spanwise lift distribution also provided advantages for the structural weight: the peak bending moment and shear of the BWB is about 50% of that for a conventional tube-and-wing configuration, allowing lighter structural weight and enhancing the structural efficiency (Liebeck, 2002).

The cross-sectional area distribution performance of the blended wing body is also worth mentioning. Fig. 32 shows a uniformly distributed cross sectional area similar to that of Sears-Haack body, which performs the lowest theoretical wave drag (Jones, 1953). Thus, the wave drag produced by the BWB is much lower than that of the conventional aircrafts (Okonkwo & Smith, 2016), and the Mach number can be further increased without change in baseline geometry of the configuration (Liebeck, 2002).

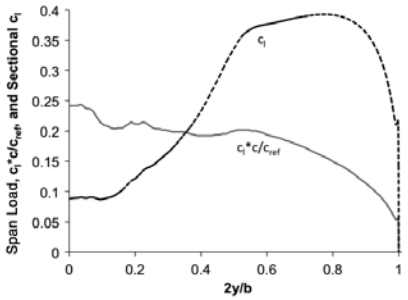


Figure 29. *Span wise lift co-efficient and span wise lift distribution of SAX-29 at 3° Angle of Attack (Sargeant et al., 2010*

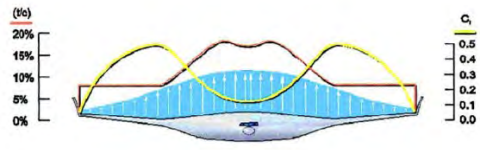


Figure 30. *Illustration graph for the Blended Wing Body: Red curve indicates Thickness to Chord ratio; Yellow curve indicates lift coefficient; Blue region indicates lift distribution (Liebeck, 2002)*

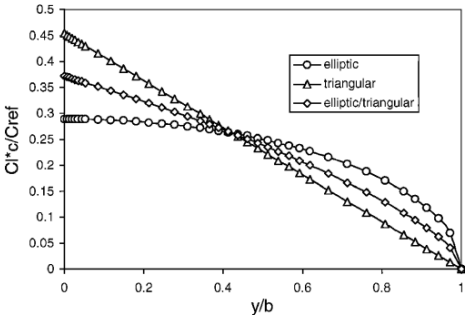


Figure 31. *Target of three different spanwise lift distribution: elliptic is shown as circle curve; triangular is shown as triangle curve and elliptic/triangular is shown as rectangular curve (Qin et al., 2005)*

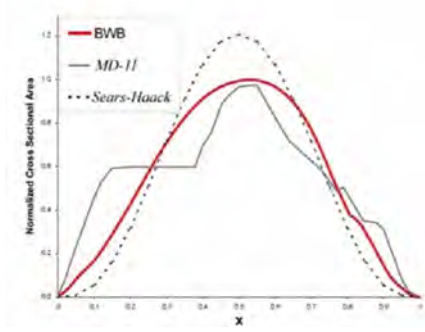


Figure 32. *Comparison of cross sectional area variation between BWB (red line), Sears-Haack body (dashed line) and MD-11 (dashed line) (Liebeck, 2002)*

5.3 New technologies and designs used in the blended wing body

Along with the novel aerodynamic configuration design, several advanced propulsion technologies have been created or used in this airframe to further improve the performance of aerodynamic efficiency and noise reduction.

5.3.1 Thrust vectoring

Thrust vectoring, referring to the ability to manipulate the direction of the thrust generated by engines, is a technology that can largely improve the mobility of the airplane and enable it to do more difficult technical movement, e.g., Cobra maneuver. It is this characteristic that makes thrust vectoring now widely used in the military field for fighters: the latest generation air fighters, Lockheed F-22A (Fig. 33), Chengdu J20 and Sukhoi Su-57, all equipped with thrust vectoring engines. However, in commercial aircraft field there is not a single airframe that utilizes this technology due to its complex design and lack of demand. The tepid response in the commercial field does not necessarily mean that there is no potential of thrust vectoring in commercial airlines. Within the whole flight cycle for the BWB aircraft, thrust vectoring is mainly used in the take-off and landing stages for pitch trim control accompanied with elevon deflection. According to Martinez-Val et al., the effect of thrust vectoring is equal to elevon deflection at 10° (Martínez-Val et al., 2007), and unlike the nature of elevon vectoring, which will unload outer wing load and increase the deck angle, thrust vectoring only requires about a 3° of Angle of Attack when climbing out (Okonkwo & Smith, 2016; Hileman et al., 2010). Thrust vectoring is favorable at the climb-out stage because it can avoid the loss in lift to drag ratio and enhance the climb-out performance (Hileman et al., 2007). However, thrust vectoring will exert extra cost on specific fuel consumption. With the extra complexity thrust vectoring adds to the aircraft, the design and manufacturing process will be burdened from time and material perspectives.

Figure 33. *The two-dimensional thrust vectoring nozzle for F-22A (2022)*

5.3.2 Boundary Layer Ingestion

Boundary Layer Ingestion (Fig. 34), by the definition provided by Hall et al., is "a propulsion concept in which some or all of the vehicle fuselage or wing boundary layers are ingested by the propulsion system and re-accelerated, instead of passing undisturbed into wakes" (Hall et al., 2017). The main benefit of boundary

layer ingestion is composed of three dimensions: enhanced propulsive efficiency due to reduction in jet dissipation, reduced wetted area and drag due to reduction in surface dissipation, and reduced wake dissipation. The related advantage is gained with a 57-69% reduction in jet dissipation, 23-38% reduction in surface dissipation and 5-8% associated with wake dissipation (Uranga et al., 2018). The advantages provided by the boundary layer ingestion can meet the design goals of a blended wing configuration: higher efficiency and lower noise. Compared to podded engines, boundary layer ingestion establishes a 23% increase in thrust and is 5-10 dB quieter (Gray et al., 2018; Romani et al., 2020). Nevertheless, the overall performance of boundary layer ingestion depends on other variables, e.g., fuselage geometry, nacelle geometry, inlet design, etc. (Gray et al., 2018). To take advantage of boundary layer ingestion, careful consideration is required between the tradeoff between propulsive efficiency and aerodynamic performance.

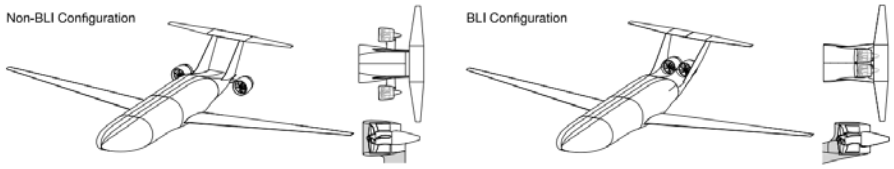


Figure 34. *Comparison of engines with non-Boundary Layer Ingestion and Boundary Layer Ingestion design on the D-8 airframe (Uranga et al., 2018)*

5.4 Problems that need to be solved

Though there are significant advantages in aerodynamic efficiency and noise reduction, there are drawbacks of the configuration and problems awaiting to be solved. In the BWB design, the use of curves makes the manufacturing process much more difficult than that of conventional aircraft. The curves and chambers in the thick center body fuselage enhance the manufacturing difficulty and increase manufacturing costs (Liebeck, 2002). BWB cabin pressure vessels also act as a great challenge in manufacturing (Liebeck, 2002). Last but not least, the current design tools cannot provide analysis that combines the physical analysis on mass estimation and aerodynamic evaluation (Liebeck, 2002). In the spanwise lift distribution research, further research requires considerations in both the spanwise lift distribution optimization and the structural weight or bending moment (Qin et al., 2005). The demand for an investigation of the balance between the lift to drag ratio (aerodynamic performance) and other parameters such as structural weight, flight stability, etc. is also mentioned (Qin et al., 2004). Currently, most of the design and analysis is still at the preliminary design stage, and deeper detailed design of electric control system or wiring is not specially designed and discussed.

6. Conclusion

This essay innovatively uses ANSYS Fluent as the investigation tool for aerodynamic evaluation of the blended wing body configuration. Besides ANSYS Fluent, a variety of resources and tools were also applied to the literature search, essay writing and data collecting. During the literature search process, the writer

learned different search methods, e.g., topic search and title search, for different purposes, and “Web of Science” and “Google Scholar” as online bases are explored and used. The airfoil modelling and aerodynamic simulation were carried out in ANSYS Fluent, and the airfoil coordinates were downloaded from the UIUC airfoil database. In ANSYS Fluent, meshing methodology was learned and evaluated, and a convergence analysis was performed. The convergence analysis helps find a reliable meshing accuracy with an acceptable computational cost. Data arrangement and visualization was completed in Wolfram Mathematica. Various functions were used to generate graphs that fit the requirement. The high operability provides flexibility in manipulating the tables and graphs drawn.

The k-ε realizable turbulence model was used for the analysis on the airfoils selected based on the design of SAX-40 airframe. An analysis based on center-body airfoil and outer-wing airfoil was carried out by using 2-D computational fluid dynamics simulation. Unfortunately, there is no airfoil similar to the center-body airfoil of SAX-40, so the analysis of the center-body airfoil is based on existing data, and the data of outer-wing airfoil is generated through ANSYS Fluent.

The baseline outer-wing section airfoil was not depicted from the origin data but selected from the UIUC airfoil data base based on the characteristics in the original design. Limited by the fidelity of the model used and the accuracy of the airfoil drawn in the CFD software, only qualitative analysis could be carried out. The simulation results generally fit the trend and description in other related papers, but the lift to drag ratio does not reach the peak figure at the cruise angle. Thus, optimization for the lift to drag ratio of NASA SC (2)-0406 airfoil as a baseline can be carried out. At cruise angle, the pressure coefficient at upper and lower surfaces near the leading edge is almost the same and most of the lift is generated near the trailing edge. Thus, the center of pressure is located near the trailing edge and creates a nose-down pitch moment. There is a positive relationship established between the Angle of Attack and lift force near the leading edge, so the center of pressure will shift forward as Angle of Attack increases.

Since the writer can neither directly access the coordinates of the SAX-40 center-body airfoil nor draw the airfoil based on existing information, the writer proposed the airfoil that is already generated and used the existing data and graphs for analysis. The pressure center of center-body airfoil is located near the leading edge due to the additional lift generated by the nose camber. The pressure distribution creates a nose-up pitch moment to the airframe, balancing the nose down moment created by the outer-wing airfoil. What is more, the pressure distribution does not have an active reaction towards the variation of Angle of Attack, showing the pitch stability of the center-body airfoil.

The 2-D simulation of airfoil is not enough for an aerodynamic evaluation of the whole airframe, and a 3-D analysis is required. Due to knowledge limitations, only a proposal can be carried out. The writer will build a 3-D model of the target airframe and use 3-D Reynolds averaged Navier-Stokes equations simulation for further analysis if possible. The 3-D analysis in this essay will use the existing graphs and data. The blended wing body is able to establish a spanwise lift distribution with the airfoil-like fuselage. Related to different cruise condition, spanwise lift distribution can be elliptic, triangular and elliptic-triangular. The blended wing body establishes a satisfying cross sectional area

distribution, indicative of the potential to further increase the cruise speed with changes in aerodynamic shape.

The blended wing body is able to take advantage of a high lift to drag ratio when cruising, and thus has a high aerodynamic and fuel efficiency. The blended wing body does not have a horizontal tail, so it largely reduces the total wetted area and drag force, endowed the airframe with high efficiency performance. By using thrust vectoring technology during the take-off climb-out, the airframe can further enhance the fuel efficiency, making it a greener airplane. Furthermore, the configuration also benefits from its outstanding noise reduction capability, and by using boundary layer ingestion, the noise reduction performance will be further improved. Boundary layer ingestion also helps reduce the energy dissipation, which also enhances aerodynamic efficiency. However, both thrust vectoring and boundary layer ingestion are not ready for commercial flight yet, and the use of new technology may bring potential risk and financial burden. The thick airfoil-like center-body and cambers at the airfoils add difficulty in the airframe manufacturing, and the multidisciplinary platform is in eager need for a further detailed design process. Market feedback will also be another challenge for this new configuration: since the first swept wing jetliner was put into service in commercial aviation, swept wing jetliners have dominated the market for decades. It is a still a question whether the market will buy this new configuration.

Conceptualized in 1988, the blended wing body configuration exhibits advanced aerodynamic performance, and the configuration has high hopes from its designers and evaluators. Until now, the conceptual design of the blended wing body is mature, preliminary design is about to complete, and detailed design of the aircraft can be carried out. However, due to the highly integrated nature of the blended wing body, multidisciplinary optimization tools need to be used for future study. Entering the 21st century, more and more countries and institutes are interested in this configuration, and McDonell Douglas, now Boeing, has been carrying out related research for more than 3 decades. In the near future, the blended wing body may enter the detailed design stage and provide a brand-new choice for the commercial aviation field.

References

Okonkwo, P., &&; Smith, H. (2016). Review of evolving trends in blended wing body aircraft design. *Progress in Aerospace Sciences*, 82, 1–23. https://doi.org/10.1016/j.paerosci.2015.12.002

Larrimer, B. I. (2020). Beyond tube-and-wing: https://www.nasa.gov/ebooks/ (Ser. NASA aeronautics book series). NASA. Retrieved 2022, from https://lccn.loc.gov/2020004750.

Dehpanah, P., &&; Nejat, A. (2015). The aerodynamic design evaluation of a blended-wing-body configuration. *Aerospace Science and Technology*, 43, 96–110. https://doi.org/10.1016/j.ast.2015.02.015

Jung Hoe, P., &&; Nik Mohd, N. A. (2014). Numerical prediction of blended wing body aerodynamic characteristics at subsonic speed. *Jurnal Teknologi*, 71(2). https://doi.org/10.11113/jt.v71.3722

Qin, N., Vavalle, A., Le Moigne, A., Laban, M., Hackett, K., &; Weinerfelt, P. (2004). Aerodynamic considerations of blended wing body aircraft. *Progress in Aerospace Sciences*, 40(6), 321–343. https://doi.org/10.1016/j.paerosci.2004.08.001

Wildschek, A., Havar, T., &; Plötner, K. (2009). An all-composite, all-electric, morphing trailing edge device for flight control on a blended-wing-body airliner. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 224(1), 1–9. https://doi.org/10.1243/09544100jaero622

Dowling, S. (2016, February 3). The WW2 flying wing decades ahead of its time. BBC Future. Retrieved July 10, 2022, from https://www.bbc.com/future/article/20160201-the-wwii-flying-wing-decades-ahead-of-its-time

Christopher, J. (2013). The race for Hitler's X-planes: Britain's 1945 mission to capture secret Luftwaffe technology. History Press.

Boyne, W. J. (1994). *Clash of Wings*. Simon &; Schuster.

Lesiv, I. (n.d.). Spotters.Aero. Retrieved August 6, 2022, from http://spotters.net.ua/

Green, W. (1972). *The Warplanes of the Third Reich*. Macdonald.

P. A., Par, -, ArnaudPassionné d'aviation tant civile que militaire depuis ma plus tendre enfance, Arnaud, &; Passionné d'aviation tant civile que militaire depuis ma plus tendre enfance. (2014, May 10). Northrop XB-35/YB-35. avionslegendaires.net. Retrieved August 6, 2022, from https://www.avionslegendaires.net/avion-militaire/northrop-xb-35-yb-35/

Cluett, N. (2022, February 17). Junkers G.38 - Germany's enormous transport. PlaneHistoria. Retrieved July 12, 2022, from https://planehistoria.com/pioneers/junkers-g-38/

Cazals, O., & Druot, T. (2013). U.S. Patent No. 8,613,409. Washington, DC: U.S. Patent and Trademark Office.

Meyers, L. M., &; Msomi, V. (2021). Hydrodynamic analysis of an underwater glider wing using ANSYS fluent as an investigation tool. Materials Today: Proceedings, 45, 5456–5461. https://doi.org/10.1016/j.matpr.2021.02.127

González, A., &; Hinojosa, J. (2019). Study of the influence of protuberances in the trailing edge of airfoils and determination of their aerodynamic efficiency through CFD using Ansys fluent. Revista Internacional De Métodos Numéricos Para Cálculo y Diseño En Ingeniería, 35. https://doi.org/10.23967/j.rimni.2019.07.001

Cayiroglu, I., &; Kilic, R. (2017). Wing aerodynamic optimization by using genetic Algoritm and Ansys. Acta Physica Polonica A, 132(3-II), 981–985. https://doi.org/10.12693/aphyspola.132.981

Wakayama, S. (2000). Blended-wing-body optimization problem setup. 8th Symposium on Multidisciplinary Analysis and Optimization. https://doi.org/10.2514/6.2000-4740

Liebeck, R. (2002). Design of the blended-wing-body subsonic transport. 40th AIAA Aerospace Sciences Meeting &; Exhibit. https://doi.org/10.2514/6.2002-2

Jones, A. R. (2006). Multidisciplinary optimization of aircraft design and takeoff operations for low noise. https://www.researchgate.net/. Massachusetts Institute of Technology. Retrieved 2022, from https://www.researchgate.net/publication/37995971_Multidisciplinary_optimization_of_aircraft_design_and_takeoff_operations_for_low_noise.

Zhonghua, H. A. N., Zhenghong, G. A. O., Wenping, S. O. N. G., &; Lu, X. I. A. (2021). On airfoil research and development: history, current status, and future directions. ACTA AERODYNAMICA SINICA, 39(6), 1–36. https://doi.org/10.7638/kqdlxxb-2021.0396

Hileman, J. I., Spakovszky, Z. S., Drela, M., Sargeant, M. A., &; Jones, A. (2010). Airframe design for silent fuel-efficient aircraft. *Journal of Aircraft*, 47(3), 956–969. https://doi.org/10.2514/1.46545

Sargeant, M. A., Hynes, T. P., Graham, W. R., Hileman, J. I., Drela, M., &; Spakovszky, Z. S. (2010). Stability of hybrid-wing-body-type aircraft with centerbody leading-edge carving. *Journal of Aircraft*, 47(3), 970–974. https://doi.org/10.2514/1.46544

Selig, M. (1996, September 11). NASA SC(2)-0406 Airfoil. UIUC airfoil data site. Retrieved August 6, 2022, from https://m-selig.ae.illinois.edu/ads/coord_database.html

Zhang, M., Chen, Z., Tan, Z., Gu, W., Li, D., Yuan, C., &; Zhang, B. (2019). Effects of stability margin and thrust specific fuel consumption constrains on multi-disciplinary optimization for blended-wing-body design. *Chinese Journal of Aeronautics*, 32(8), 1847–1859. https://doi.org/10.1016/j.cja.2019.05.018

Qin, N., Vavalle, A., &; Le Moigne, A. (2005). Spanwise lift distribution for blended wing body aircraft. *Journal of Aircraft*, 42(2), 356–365. https://doi.org/10.2514/1.4229

Messerschmitt Me 323. Military Wiki. (n.d.). Retrieved 2022, from https://military-history.fandom.com/wiki/Messerschmitt_Me_323

Qin, N. (2002). Aerodynamic studies for Blended Wing Body Aircraft. 9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization. https://doi.org/10.2514/6.2002-5448

Jones, R. T. (1953). Theory of wing-body drag at supersonic speeds (No. NACA-RM-A53H18a).

Martínez-Val, R., Pérez, E., Alfaro, P., &; Pérez, J. (2007). Conceptual design of a medium size flying wing. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 221(1), 57–66. https://doi.org/10.1243/09544100jaero90

Hileman, J., Spakovszky, Z., Drela, M., &; Sargeant, M. (2007). Airframe design for "Silent aircraft". 45th AIAA Aerospace Sciences Meeting and Exhibit. https://doi.org/10.2514/6.2007-453

Hall, D. K., Huang, A. C., Uranga, A., Greitzer, E. M., Drela, M., &; Sato, S. (2017). Boundary layer ingestion propulsion benefit for transport aircraft. *Journal of Propulsion and Power*, 33(5), 1118–1129. https://doi.org/10.2514/1.b36321

Uranga, A., Drela, M., Hall, D. K., &; Greitzer, E. M. (2018). Analysis of the aerodynamic benefit from boundary layer ingestion for transport aircraft. *AIAA Journal*, 56(11), 4271–4281. https://doi.org/10.2514/1.j056781

Gray, J. S., Mader, C. A., Kenway, G. K., &; Martins, J. R. (2018). Modeling boundary layer ingestion using a coupled aeropropulsive analysis. *Journal of Aircraft,* 55(3), 1191–1199. https://doi.org/10.2514/1.c034601

Romani, G., Ye, Q., Avallone, F., Ragni, D., &; Casalino, D. (2020). Numerical Analysis of Fan Noise for the nova boundary-layer ingestion configuration. *Aerospace Science and Technology*, 96, 105532. https://doi.org/10.1016/j.ast.2019.105532

Kundu, P. K., Cohen, I. M., &; Dowling, D. R. (2016). Fluid mechanics. Elsevier.

Aprovitola, A., Aurisicchio, F., Di Nuzzo, P. E., Pezzella, G., &; Viviani, A. (2022). Low speed aerodynamic analysis of the N2A hybrid wing–body. *Aerospace*, 9(2), 89. https://doi.org/10.3390/aerospace9020089

Li, P., Zhang, B., Chen, Y., Yuan, C., &; Lin, Y. (2012). Aerodynamic design methodology for blended Wing Body Transport. *Chinese Journal of Aeronautics*, 25(4), 508–516. https://doi.org/10.1016/s1000-9361(11)60414-7

Okonkwo, P. (2016). Conceptual Design Methodology for Blended Wing Body Aircraft. ResearchGate. Retrieved 2022, from https://www.researchgate.net/publication/357657519_Conceptual_Design_Methodology_for_Blended_Wing_Body_Aircraft.

Rahimi, H., Medjroubi, W., Stoevesandt, B., &; Peinke, J. (2014). 2D numerical investigation of the laminar and turbulent flow over different airfoils using openfoam. *Journal of Physics: Conference Series*, 555, 012070. https://doi.org/10.1088/1742-6596/555/1/012070

Why doesn't the F-22 have roll thrust vectoring? Quora. (n.d.). Retrieved August 6, 2022, from https://www.quora.com/Why-doesn-t-the-F-22-have-roll-thrust-vectoring

Leone, D. (2019, August 21). Why Boeing's B-47 Stratojet bomber was a game changer for the Air Force. *The National Interest*. Retrieved August 8, 2022, from https://nationalinterest.org/blog/buzz/why-boeings-b-47-stratojet-bomber-was-game-changer-air-force-75131

Nonea, V., Iacob, R., &; Rebedea, T. (2021). Reinforcement learning agent for a flight simulation video game. RoCHI - International Conference on Human-Computer Interaction. https://doi.org/10.37789/rochi.2021.1.1.15

A Review of Non-classic Biomanipulation Experiments at Freshwater Lakes in China and the Factors that Influence their Results

Xiaohan Zhang

Author Background: *Xiaohan Zhang grew up in China and currently attends Wuhan Britain-China School in Wuhan, Hubei, in China. Her Pioneer research concentration was in the field of environmental studies and titled "Water Quality and Global Environmental Health."*

Abstract

The toxic cyanobacterial bloom is becoming a challenge to freshwater lakes in China. To mitigate this threat, biomanipulation is one intervention that has been proposed. This study reviewed 6 non-traditional biomanipulation experiments from the past 20 years at 4 freshwater lakes in China: Lake Taihu in Jiangsu province, Lake Erhai in Yunnan province, Lake Donghu in Wuhan city, and Lake Shichahai in Beijing city.Here, non-traditional biomanipulation referred to the intentional introduction of planktivorous species such as silver carp and bighead carp to reduce cyanobacteria occurrence. We identified non-traditional biomanipulation experiments that made use of fish enclosures and calculated summary statistics according to the data presented in the result section of each study. Multiple regression analysis was done to investigate the influences of factors either relevant to or independent of biomanipulation effects on the growth of cyanobacteria and the relative size of the impact of these influencing factors. It was found that within Lake Taihu, the rate of cyanobacteria growth was negatively associated with pH in the biomanipulation enclosure, and pH had the greatest impact among all variables, with a correlation coefficient of -260. Cyanobacteria density was inversely associated with silver carp biomass per m^3 (Correlation coefficient magnitude 0.285), while bighead carp biomass density was positively associated with cyanobacteria biomass (Correlation coefficient magnitude 16.2). It was clear that silver carp exerted negative effects on cyanobacterial growth, while bighead carp exerted positive effects. Moreover, the effect size of silver carp on cyanobacterial blooms was smaller than that of bighead carp. This indicated that the silver carp was preferred over bighead carp in biomanipulation, and using silver carp only was preferred over using a combination of silver carp and bighead carp. In the comparison between different lakes, the mean depth of the lake, the most impactful variable, was positively correlated with the growth of cyanobacteria, with a correlation coefficient of 54.5. Meanwhile, silver carp biomass (correlation coefficient -0.0393), though negatively

correlated with cyanobacterial growth, had the least impact, suggesting that the current stocking densities were not sufficient to effectively control algal blooms. Through the comparison of correlation coefficient magnitude (0.285 VS 0.0393), our analysis also indicated that increasing silver carp biomass density was more effective at reducing cyanobacterial blooms within one lake than across different lakes.

1. Introduction

Freshwater lakes play an important role in people's lives. They are sources of drinking water and places for recreational activities. Meanwhile, these lakes have always been vulnerable to pollution. With the expansion of urban settlements, industrial activities, and agriculture, water pollution of freshwater lakes has become more and more severe such that it poses an increasing threat to the health of nearby residents. One major water quality threat is eutrophication caused by nutrient loads from nearby agricultural and industrial areas.

Eutrophication is the gradual increase in the concentration of plant nutrients in aquatic ecosystems, such as freshwater lakes ("Eutrophication | Definition, Types, Causes, & Effects," n.d.). Phosphorus and nitrogen are two important limiting nutrients for the growth of phytoplankton in freshwater, with phosphorus often being the primary driver (Chorus & Welker, 2021). Both nitrogen and phosphorus enter water bodies as run-off from animal feedlots, sewage, and soils, particularly if the soils were fertilized with minerals or manure (Chorus & Welker, 2021).

Eutrophication is a serious environmental concern because it often results in the depletion of dissolved oxygen in water bodies by aerobic bacteria ("EUTROPHICATION," n.d.). Such eutrophic waters can eventually become "dead zones" that are incapable of supporting any aquatic life ("EUTROPHICATION," n.d.).

Cyanobacteria, commonly called blue-green algae, are photosynthetic bacteria that inhabit nearly all surface waters (Bartram, 2015). Cyanobacteria often dominate plankton communities seasonally or perennially, depending on the temperature zone in eutrophic lakes, and can produce large surface "blooms" (Bartram, 2015). Excessive growth of cyanobacteria and surface bloom formations are completely natural phenomena in surface waters located in nutrient-rich basins (Bartram, 2015). However, increased nutrient loading of surface waters resulting from human development and activities, or anthropogenic eutrophication, is contributing to an increase in the occurrence and severity of blooms globally (Bartram, 2015).

Cyanobacterial blooms can cause numerous issues, impeding recreation by giving off offensive odors as well as reducing water clarity and hindering drinking water use both through the production of non-toxic odorous substances and, more importantly, potent hepato- (liver) and neuro-toxins (Bartram, 2015). In fact, "cyanotoxins are among the most toxic naturally occurring compounds" and present great potential risks to humans as well as domestic and wild animals (Bartram, 2015; Chorus & Welker, 2021). Also, algal blooms have the potential to deplete dissolved oxygen in stratified lakes, particularly, where internal water exchange is not prominent, contributing to fish kills. This in turn leads to the loss of economic benefits for fish farming or the propagation of water-borne bacteria.

Eutrophication that results in pervasive cyanobacterial blooms is becoming a challenge in several freshwater lakes in China. One typical example of such a lake is Lake Taihu, China's third largest freshwater lake (Qin et al., 2019). This shallow lake located in a large, agricultural catchment in the Yangtze River Delta region has experienced accelerated eutrophication accompanied by toxic cyanobacterial blooms since the 1990s (Qin et al., 2019). In May 2007, "a massive bloom overwhelmed the drinking water plants" at the lake, leading to a water crisis in Wuxi, Jiangsu, leaving millions of residents without potable water for almost a week (Qin et al., 2019).

To control cyanobacteria blooms, biomanipulation is often used as a management tool. Biomanipulation, a term first introduced in 1975, alters phytoplankton community composition and growth by influencing parts of the food web of a lake (Chorus & Welker, 2021; "Biomanipulation," 2005). One common approach in biomanipulation involves relieving the predation pressure on the zooplankton community, which in turn feeds on phytoplankton, through the removal of planktivorous fish (Yin, Guo, Yi, Luo, & Ni, 2017). Such an approach is termed "classic biomanipulation ". However, such classic biomanipulation usually malfunctions in hypertrophic lakes as it may result in the dominance of grazing-resistant "phytoplankton species such as colony-forming (*Microcystis, Aphanizomenon)* or filamentous cyanobacteria (*Planktothrix agardhii*)"(Chorus & Welker, 2021; Yin, Guo, Yi, Luo, & Ni, 2017).

A non-classic biomanipulation method was developed to tackle the issue of cyanobacterial growth in nutrient-rich lakes and has been reported to be effective over the past 30 years (Chorus & Welker, 2021). Successful cases include Lake Donghu and Lake Qiandaohu in China, where long-term stocking of silver carp was carried out (Yi et al., 2016). Non-classic biomanipulation involves the introduction of filter-feeding planktivorous fish, usually silver carp and bighead carp, that "directly collect food by filtering water via their gill rakers and hence, unselectively ingest plankton and detritus" (Yin, Guo, Yi, Luo, & Ni, 2017). However, when such a method was applied to more lakes in China, the effects varied significantly. While some sites of study achieved effective control of algae blooms, an increase in cyanobacteria growth and nutrient concentration in the lake was observed at other sites.

Over the past 20 years, though various experiments to investigate the effect of non-classic biomanipulation on cyanobacterial bloom have been done, these cases were all reported separately. Therefore, it would be hard to gain an integrated view of how the implementation of biomanipulation in lakes with different environmental conditions influences cyanobacterial blooms. This paper is the first study to review and synthesize learning from previous individual studies to learn about factors either relevant to or independent of biomanipulation effects in shallow freshwater lakes in China and their relative importance through multiple regression analysis.

Such research can be used as a reference for future policymakers and can help them implement biomanipulation measures according to the conditions of the lake.

2. Method

To start with, we identified non-traditional biomanipulation experiments using silver carp and, in some studies, bighead carp as well at freshwater lakes of different characteristics (depth, surface area, trophic state, and more) in China. All experiments studied in this paper were carried out in biomanipulation enclosures. Certain densities of fish were stocked into the enclosures at the beginning of each study. Among the lakes, Lake Taihu is the most extensively studied one, with the largest number of papers available for review. Therefore, in addition to comparing different lakes, a comparison was also made between 3 studies within Lake Taihu carried out at its two bays to better illustrate the effect of different fish stocking types and densities on the growth of cyanobacteria. Originally, attempts were made to collect raw direct measurement data from the Chinese Ecosystem Research Network. However, due to the database's restriction that the data obtained cannot be translated into languages other than Chinese, they became unavailable for use in this paper. Thus, all the direct measurement data that appeared in this study came from the result sections of the 6 papers reviewed.

2.1 Study Site

2.1.1 Lake Taihu

Located downstream of the Yangtze River, Lake Taihu (30°55′-31°33′ N, 119°52′-120°36′ E) "is approximately 2338 km^2 with a mean depth of 1.89 m and a maximum depth of 2.6 m" (Yi et al., 2016; Qin et al., 2007).

2.1.1.1 Meiliang Bay of Lake Taihu

Meiliang Bay is one of the most eutrophic bays in the northeastern part of Lake Taihu, with a depth of 1.8-2.3 m and a surface area of 100 km^2 (Ye et al., 2015). Two main inflowing rivers, the Zhihugang and the Liangxi, connect to Meiliang Bay (Ye et al., 2015). Untreated wastewater heavily polluted by industrial and agricultural activities, as well as domestic sewage from residential areas and factories, is discharged into both rivers, leading to frequent *Microcystis* spp. blooms in Meiliang Bay in the past decades (Ye et al., 2015).

2.1.1.2 Gonghu Bay of Lake Taihu

"Gonghu Bay is in the north-east section of Taihu Lake and has a surface area of 150 km^2 and a depth of 1.8–2.5 m, which is the main source of drinking water for Wuxi City" (Guo et al., 2014). "Because Gonghu Bay relies on Wangyu River to direct water from the Yangtze River into the lake, Wangyu River could significantly affect the sources and concentrations of nitrogen and phosphorus pollutants" (Guo et al., 2014).

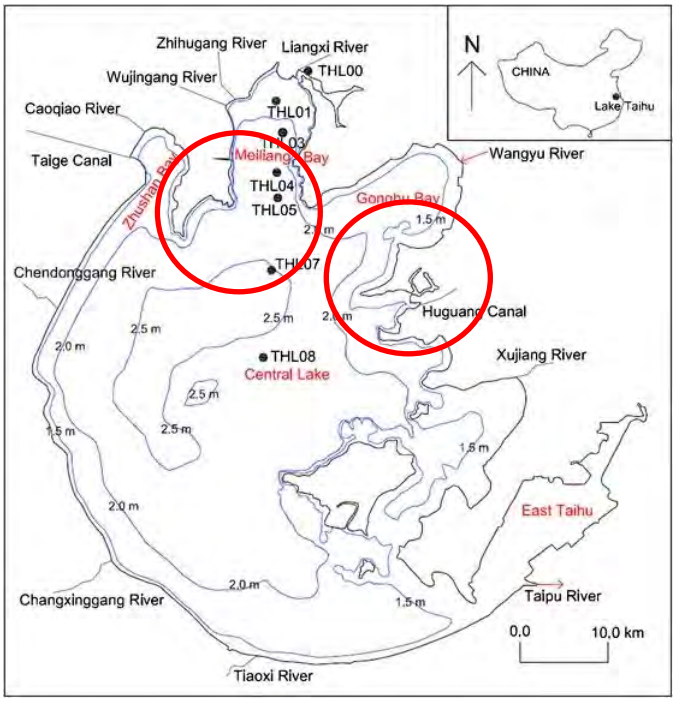


Diagram 1. *The study sites at Lake Taihu (Akyuz, Luo & Hamilton, 2014)*

2.1.2 Lake Erhai

Lake Erhai (25°39′-25°56′ N, 100°06′-100°17′ E) is a plateau lake located at 1973m altitude in Yunnan Province and has a surface area of 249 km^2 (Yi et al., 2016). "The mean and maximum depths of Lake Erhai are approximately 10.7 m and 22 m, respectively" (Yi et al., 2016).

2.1.3 Lake Donghu

Lake Donghu (30° 33′, 114°23′E) is a 28 km^2 shallow (mean depth 2.2 m, maximum depth 4.8m), subtropical lake only five kilometers away from the Yangtze River (P.R. China) ("Lake Donghu | Donghu | World Lake Database - ILEC," n.d.). Its location is at the eastern end of Wuhan City, Hubei Province ("Lake Donghu | Donghu| World Lake Database - ILEC," n.d.).

2.1.4 Lake Shichahai

"Lake Shichahai (39°58′N, 116°29′E) in Beijing City is a shallow hypereutrophic temperate lake with a surface area of 17.9 km^2, and a mean depth of 1.3 m" (Zhang et al., 2006).

2.2 Experimental Data

2.2.1 2004-2005 Experiment at Meiliang Bay, Lake Taihu[1]

Three fish pens were built for stocking silver and bighead carp in Meiliang Bay. Each

[1] All methods, diagram presented below and numerical data used in the later analysis were from the paper of Ke et al. (2007).

fish pen had an area of 0.36 km^2 , and the water depth was around 2m on average. The "mesh size of the net was 2 cm × 2 cm".

24,775 kg silver carps and 8005 kg bighead carps were stocked on average into the three pens from December 2004 to January 2005. Thus, the initial stocking densities of silver carp and bighead carp were 11.47 g/m^3 and 3.71 g/m^3, respectively. The sampling of fish, as well as zooplankton and phytoplankton, was done monthly in 2005.

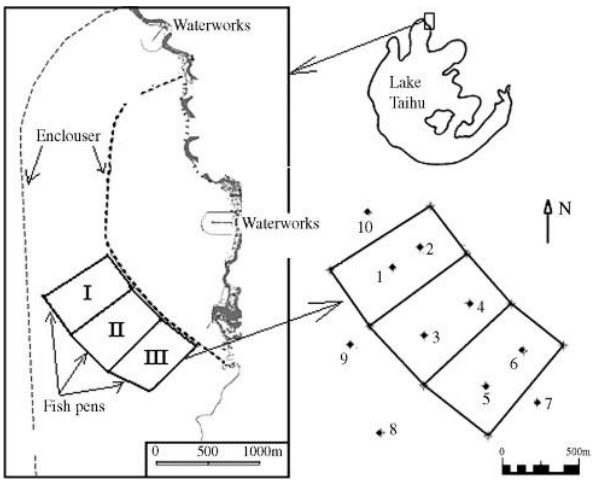


Diagram 2. *The rough location of the sampling sites in Meiliang Bay*

For zooplankton and phytoplankton sampling, water samples were collected from the ten sampling sites shown in Diagram 2 (six in the pens and four in the surrounding water).

For sampling from the ten sampling sites, using a Patalas–Schindler trap, discrete water samples from the depth of 0m, 0.5m, 1.0m, and 1.5 m were taken respectively at each sampling site.

Wet weights of crustacean zooplankton and phytoplankton were estimated and physicochemical parameters including water temperature, total phosphorus (TP), total nitrogen (TN), pH, and transparency were assessed according to standard methods (as described in Ke et al. (2007)). Their annual means were calculated.

2.2.2 2009 Gonghu Bay Experiment, Lake Taihu[2]

For this experiment, there were four sampling sites: Station 1 and 2 were set up in the center of a fish enclosure with a total area of 0.08 km^2 (length 320 m, width 250 m). Station 3 and 4 were set up in the surrounding lake as a control. There was no water exchange between the surroundings and the enclosure. The initial density for silver carp was 7.5 g m^{-3} and the initial density for bighead carp was 1.1 g m^{-3} . The mass ratio of silver carp to bighead carp was 4:1. The fish were stocked into the lake in March 2009, and the experiment lasted throughout 2009.

[2] All methods, diagram, and numerical data used in the later analysis were from the paper of Guo et al. (2015).

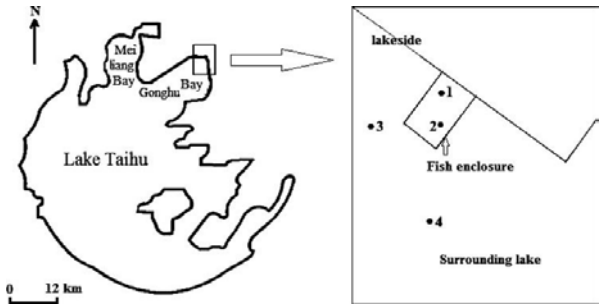


Diagram 3. *Rough location of the sampling sites in Gonghu Bay*

To measure the water parameters, a Patalas-Schindler trap was used to collect monthly integrated water column samples in 2009. The parameters measured and the measuring methods are shown below:

Table 1. *Parameters measured and the measuring methods*

Parameters	Method of measuring
Water temperature	"YSI Environmental Monitoring System 6600 (YSI Incorporated, Yellow Springs, OH, USA)"
pH	
Transparency	Secchi disk
Total nitrogen (TN)	Absorbance measured as nitrate at 220 nm after digestion of the total samples with $K_2S_2O_8$ +NaOH
Total phosphorus (TP)	Colorimetry carried out after digestion of the total samples to orthophosphate with $K_2S_2O_8$ +NaOH
Microcystis biomass	Detailed method described in Guo et al.(2015)
Crustacean zooplankton biomass	

2.2.3 2011 Experiment at Meiliang Bay, Lake Taihu[3]

This experiment continued from 30 May to 23 June 2011, sampling at an interval of 3-4 days. Four identical cuboid waterproof PVC enclosures (2.5m × 2.5m × 3 m) were set up in Meiliang Bay, Lake Taihu, placed at a depth of 1.5~ 1.6m approximately within the littoral zone. Silver carp were collected from a local fish farm and acclimated in a nearby pond for ten days prior to being put into the enclosures. For the total of 24 enclosures, four biomass treatments were performed in triplicate: 0, 35, 70, and 150 g m^{-3} . Among the 24 enclosures, 12 enclosures were then chosen and divided randomly into four groups: control group (CG), low fish density group (LDG), medium fish density group (MDG), and high fish density group (HDG). The biomass of silver carp for each group was the following:

[3] All methods and numerical data used in the later analysis were from the papers of Yi et al. (2016) and Yin et al. (2017).

Table 2. *Silver carp biomass*

Group	Fish Biomass treatment (g/m^3)
CG	0
LDG	35
MDG	70
HDG	150

Integrated water samples were collected with a 5 L modified Patalas's bottle sampler. A sample with a total volume of 10 L was collected: 5 L from 0.5 m below the surface of the water and 5 L from 0.5 m above the bed of the lake.

The parameters measured and the measuring method used were summarized in the table below:

Table 3. *Parameter measured and measuring methods used*

Parameters	Method of Measuring
pH	"YSI professional plus water quality monitor (YSI Inc., Yellow Springs, Ohio, USA)"
Water temperature	
Transparency (Zsd)	20-cm in diameter Secchi's disk
Total nitrogen (TN)	"Standard Methods for the Examination of Water and Wastewater"
Total phosphorus (TP)	
Microcystis biomass	Detailed method described in Yin et al. (2017)
Crustacean zooplankton biomass	

Also, for the 12 enclosures chosen to set up CG, LDG, MDG, and HDG, quantitative measurement of *Microcystis* spp. and crustacean zooplankton biomass was carried out.

2.2.4 2014 Lake Erhai Experiment[4]

All the methods used and parameters measured in this experiment are the same as those in the 2011 Meiliang Bay, Lake Taihu experiment, except for the silver carp biomass being 0, 20, 50, and 100 g m^{-3}. Samplings at the enclosures were done weekly from 22 August 2014 to 24 October 2014.

2.2.5 1999 Lake Donghu Experiment[5]

The experiment lasted from 29 April 1999 until 26 June 1999. Eight polyethylene enclosures (2.5 × 2.5 m by 2 m depth) were set up at Lake Donghu. "Silver carp were collected from a nearby pond." They were acclimated in a net cage placed in the lake for several days and then transferred into the enclosures. Four fish biomass (two replicates each) were tested:

[4] All methods and numerical data used in the later analysis were from the paper of Yi et al. (2016).

[5] All methods and numerical data used in the later analysis were from the paper of Tang et al. (2002).

Table 4. *Fish biomass*

Treatment	Biomass of Fish (g/m^3)
No fish	0
Low biomass	58
Median biomass	88
High biomass	158

Here, the unit for fish biomass was converted from g/m^2 in the original paper to g/m^3

Transparency and temperature measurements and sampling for water chemistry, crustacean zooplankton, and phytoplankton were done weekly during the study period. Parameters measured are shown below:

Table 5. *Parameters measured and methods of measuring*

Parameters	Method of Measuring
Transparency	Secchi disc
Total nitrogen (TN)	Absorption measured as nitrate at 220 nm after digestion with alkaline potassium persulfate under 120 °C for half an hour
Total phosphorus (TP)	Measured as orthophosphate at 220 nm and analyzed with ammonium molybdate method after digestion with alkaline potassium persulfate under 120 °C for half an hour
Phytoplankton biomass	Detailed method described in Tang et al. (2002)

2.2.6 2004 Lake Shichahai Experiment[6]

"The experiment lasted from June 15 to October 25, 2004." Eight polyhexene enclosures (3 × 3 × 2.5 m) were set up "in the most eutrophic part of Lake Shichahai," each "filled with lake water to a depth of 1 m." "Silver carp were collected from a nearby pond and acclimated in a net cage" in the lake before being put into the enclosures on June 15. Four fish biomass levels were chosen (shown below), each with two replicates.

Table 6. *Treatments and the corresponding fish biomass*

Treatment	Biomass of Fish (g/m^3)
No fish	0
Low biomass	58
Median biomass	88
High biomass	158

Mixed water samples were taken with a 5-l modified Patalas's bottle sampler from the surface and bottom layers of the lake. Parameters measured were shown as below:

[6] All methods and numerical data used in the later analysis were from the paper of Zhang et al. (2006)

Table 7. *Parameters and the corresponding measuring methods*

Parameters	Method of measuring
Transparency	Secchi disc
Total nitrogen (TN)	Absorbance measured as nitrate at 220 nm after digestion of the unfiltered samples with $K_2S_2O_8$ + NaOH
Total phosphorus (TP)	Colorimetry carried out after digestion of the unfiltered samples to orthophosphate with $K_2S_2O_8$ + NaOH
Phytoplankton biomass	Detailed method described in Zhang et al. (2006)

2.2.7 Weather Data

For each experiment, the average precipitation of the nearest major city during the months that the study period mainly covered in the corresponding year was collected from the *China Statistical Yearbook*. Here, for a month to be mainly covered by the study period, the study period needed to include a least 15 days of that month. For Lake Taihu, precipitation of Shanghai was used; For Lake Erhai, precipitation of Kunming was used; For Lake Donghu, precipitation of Wuhan was used; For Lake Shichahai, precipitation of Beijing was used.

2.3 Data analysis

For data analysis, multiple regression was carried out through Excel.

Three experiments in Lake Taihu were compared: 2005 Meiliang Bay, 2011 Meiliang Bay, and 2009 Gonghu Bay. The independent and dependent variables used are shown as the following:

Table 8. *Independent and dependent variables used*

Independent Variable	Dependent Variable
Silver carp biomass (g/m^3)	Percentage decrease in mean *Microcystis* or total phytoplankton biomass compared to the control
Bighead carp biomass (g/m^3)	
Mean pH	
Mean water temperature (℃)	
Mean crustacean zooplankton biomass (μg/L)	
Mean transparency (cm)	
Precipitation (mm)	

Microcystis biomass and total phytoplankton biomass have a close correlation as the phytoplankton consists mainly of *Microcystis* during the blooms. Therefore, mean *Microcystis* biomass and mean total phytoplankton biomass were used alternatively to reflect the growth of cyanobacteria. Mean transparency was used as an independent variable to reflect light availability.

Four experiments: 2011 Lake Taihu Meiliang Bay, 2014 Lake Erhai,1999 Lake Donghu, and 2004 Lake Shichahai, were compared. The independent and dependent variables used are shown as the following:

Table 9. *Independent and dependent variables used*

Independent variable	Dependent variable
Silver carp biomass (g/m^3)	Percentage change in mean chlorophyll-a concentration or phytoplankton biomass compared to the control
Mean depth of the lake (m)	
Mean water temperature (℃)	
Precipitation (mm)	
Mean lake TP (mg/L)	
Mean lake TN (mg/L)	
Surface area of the lake (km^2)	
Mean transparency (cm)	
Silver carp biomass (g/m^3)	Percentage change in mean total nitrogen compared to the control
Mean depth of the lake(m)	
Mean water temperature (℃)	
Precipitation(mm)	
Mean lake TP (mg/L)	
Mean lake TN (mg/L)	
Surface area of the lake (km^2)	
Silver carp biomass(g/m^3)	Percentage change in mean total phosphorus compared to the control
Mean depth of the lake(m)	
Mean water temperature (℃)	
Precipitation(mm)	
Mean lake TP (mg/L)	
Mean lake TN (mg/L)	
Surface area of the lake (km^2)	

Here, mean lake total phosphorus and mean lake total nitrogen were used to reflect the trophic state of the lakes.

Table 10. *Trophic state categories and their definition in terms of mean total phosphorus as given by Vollenweider & Kerekes (1982)*

Trophic state	TP mean (µg/L)
Ultraoligotrophic	≤ 4
Oligotrophic	≤ 10
Mesotrophic	10 – 35
Eutrophic	35 – 100
Hypertrophic	≥100

According to this standard, the trophic state classification for the four lakes could be shown as the following:

Table 11. *Trophic state classification*

	Mean lake TP (µg/L)	Trophic state
Lake Taihu	800	Hypertrophic
Lake Erhai	30	Mesotrophic
Lake Donghu	570	Hypertrophic
Lake Shichahai	230	Hypertrophic

Since "chlorophyll-a is a widely used and accepted measure" of total phytoplankton biomass (Chorus & Welker, 2021), chlorophyll-a concentration and phytoplankton biomass went together to reflect the growth of cyanobacteria. When the dependent variable was the percentage change in mean chlorophyll-a concentration or phytoplankton biomass, mean transparency was used as an independent variable to reflect light availability.

3.Result

3.1 Regression Analysis for Comparison between Different Biomanipulation Experiments at Lake Taihu

Table 12. *Regression coefficient for the independent variables when percentage decrease in mean Microcystis/total phytoplankton biomass was the dependent variable.*

Independent Variable	Coefficient (3 significant figures)
Silver carp biomass	0.285
Bighead carp biomass	-16.2
Mean pH	-260
Water temperature	-90.0
Mean crustacean zooplankton biomass	-84.3
Mean transparency	-0.323
Precipitation	1.92

The regression analysis showed that the percentage decrease in mean *Microcystis* or total phytoplankton biomass compared to the control was positively correlated with silver carp biomass and precipitation, while negatively correlated with mean pH, mean water temperature, mean crustacean zooplankton biomass, mean transparency, and bighead carp biomass. In terms of their influence on the dependent variable from the greatest to the smallest, the independent variables could be listed in the order: mean pH, mean water temperature, mean crustacean zooplankton biomass, bighead carp biomass, precipitation, mean transparency, and silver carp biomass. The R squared value for this regression was 1, which meant that the independent variables and the dependent variable had a perfect linear relationship.

3.2 Regression Analysis for Comparison between Biomanipulation Experiments at Different Lakes

Table 13. *Percentage change in mean transparency in the fish treatments compared to the control*

Study site	Low biomass	Median biomass	High biomass
Lake Taihu	21.3	9.46	-4.51
Lake Erhai	-32.0	-30.0	-20.7
Lake Donghu	-53.7	-37.3	-32.2
Lake Shichahai	-49.0	-17.8	-41.0

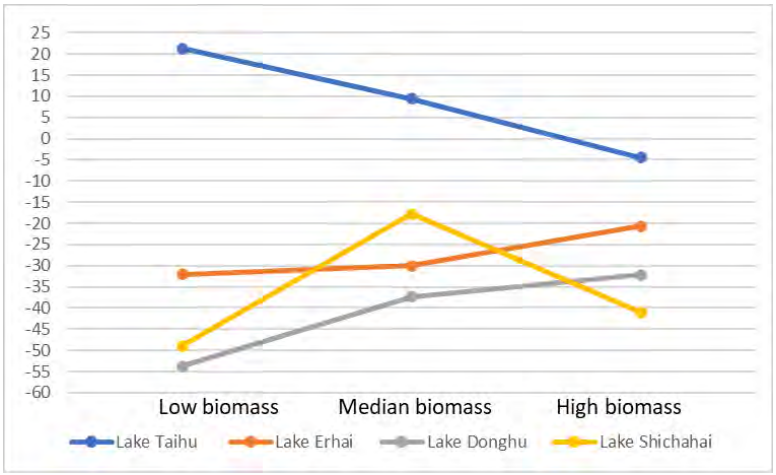


Figure 1. *Percentage change in mean transparency in the fish treatments compared to the control*

Table 14. *Percentage change in mean total nitrogen in the fish treatments compared to the control*

Study site	Low biomass	Median biomass	High biomass
Lake Taihu	-23.1	-0.641	-17.3
Lake Erhai	-7.41	-1.23	-1.23
Lake Donghu	-11.1	-2.02	-16
Lake Shichahai	54.5	2.73	67.3

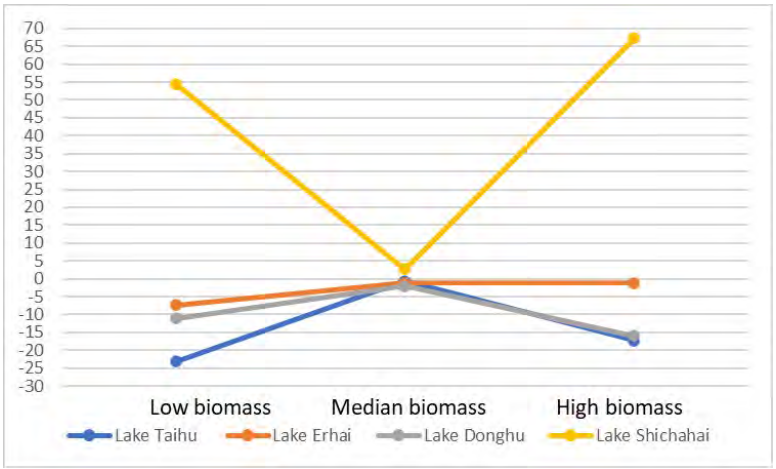


Figure 2. *Percentage change in mean total nitrogen in the fish treatments compared to the control*

Table 15. *Percentage change in total phosphorus in the fish treatments compared to the control*

Study site	Low biomass	Median biomass	High biomass
Lake Taihu	-40.0	-55.0	-50.0
Lake Erhai	6.90	17.2	-6.90
Lake Donghu	-26.7	-6.67	-28.3
Lake Shichahai	100	1000	129

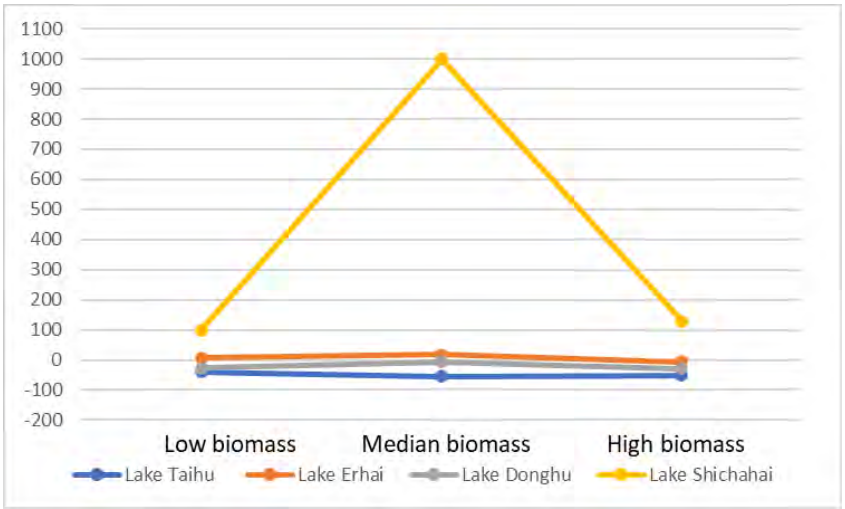


Figure 3. *Percentage change in total phosphorus in the fish treatments compared to the control*

Table 16. *Percentage change in phytoplankton biomass in the fish treatments compared to the control*

Study site	Low biomass	Median biomass	High biomass
Lake Donghu	597	570	356
Lake Shichahai	97.8	-46.1	157

Table 17. *Percentage change in chlorophyll-a concentration in the fish treatments compared to the control*

Study site	Low biomass	Median biomass	High biomass
Lake Taihu	-69.9	-58.2	-69.4
Lake Erhai	74.4	75.7	62. 4

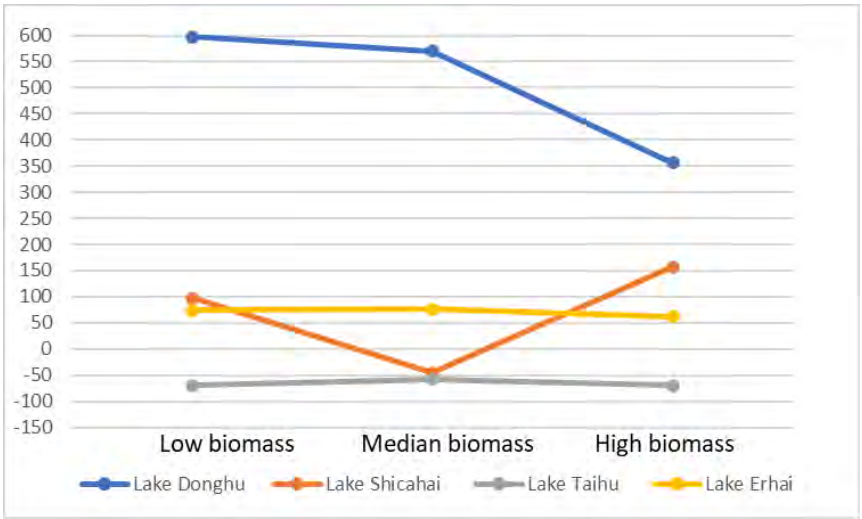


Figure 4. *Percentage change in mean chlorophyll-a concentration or phytoplankton biomass in the fish treatments compared to the control*

Table 18. *Regression coefficient for the independent variables when percentage decrease in mean chlorophyll-a or phytoplankton biomass was the dependent variable*

Independent Variable	Coefficient (3 significant figures)
Biomass of silver carp	-0.0393
Mean depth of the lake	54.5
Mean water temperature	0
Precipitation	2.38
Mean Lake TP	0
Mean Lake TN	0
Surface area of the lake	-0.263
Mean Transparency	-4.60

The percentage change in mean chlorophyll-a concentration or phytoplankton biomass compared to the control was negatively correlated with silver carp biomass, mean transparency, and surface area of the lake, while positively correlated with the mean depth of the lake and precipitation. Also, there was no correlation between the percentage change in mean chlorophyll-a concentration or phytoplankton biomass and mean lake TN, TP (trophic state) or mean water temperature.

Among the independent variables that were correlated with the dependent variable, in terms of their influence on the dependent variable from the greatest to the smallest, these independent variables could be listed in the order: mean depth of the lake, mean transparency, precipitation, surface area of the lake, and the biomass of silver carp. The R squared value for this regression was 0.877 (3 significant figures), which meant that 87.7% of the variance in the dependent variable could be explained by the independent variables.

Table 19. *Regression coefficient for the independent variables when the percentage change in mean total nitrogen was the dependent variable*

Independent Variable	Coefficient (3 significant figures)
Biomass of silver carp	0.0251
Mean depth of the lake	-3.33
Water temperature	0
Precipitation	-0.147
Mean Lake TP	0
Mean Lake TN	0
Surface area of the lake	-0.00181

Table 20. *Regression coefficient for the independent variables when the percentage change in mean total phosphorus was the dependent variable*

Independent Variable	Coefficient (3 significant figures)
Biomass of silver carp	0.234
Mean depth of the lake	-30.2
Mean water temperature	0
Precipitation	-1.23
Mean Lake TP	0
Mean Lake TN	0
Surface area of the lake	-0.0136

Percentage change in total nitrogen and total phosphorus compared to the control shared the same trend in relation to the independent variables. The percentage change was positively correlated with silver carp biomass, while negatively correlated with the mean depth of the lake, precipitation, and lake surface area. There was no impact of mean water temperature or trophic level of the lake (mean lake TP and TN) on the change in total nitrogen and total phosphorus.

Among the independent variables that were correlated with the dependent variable, in terms of their influence on the dependent variable from the greatest to the smallest, these independent variables could be listed in the order: mean depth of the lake, precipitation, biomass of silver carp, and the surface area of the lake. For the regression analysis with the percentage change in total nitrogen as the dependent variable, the R squared value was 0.512 (3 significant figures), which meant that 51.2% of the variance in the dependent variable could be explained by the independent variables. For the regression analysis with the percentage change in total phosphorus as the dependent variable, the R squared value was 0.329 (3 significant figures), which meant that 32.9% of the variance in the dependent variable could be explained by the independent variables.

4.Discussion

4.1 Comparison between Different Biomanipulation Experiments at Lake Taihu

“Silver carp mainly fed on phytoplankton while bighead carp mainly fed on

zooplankton," another predator of cyanobacteria (Ke, Xie, Guo, Liu, & Yang, 2007). That explained why mean *Microcystis* or total phytoplankton biomass decreased more as silver carp biomass increased, but decreased less as bighead carp biomass increased.

Mean *Microcystis* or total phytoplankton biomass decreased less as mean crustacean zooplankton biomass increased. Silver carp and bighead carp feed on both crustacean zooplankton and phytoplankton (Ke, Xie, Guo, Liu, & Yang, 2007). However, according to Ke, Xie, Guo, Liu, & Yang (2007), "there was a greater proportion of zooplankton in the guts of silver and bighead carp" compared to phytoplankton. This suggests that when zooplankton became more and more abundant, the diet of silver carp and bighead tended to shift toward zooplankton, which was a food source with more energy than phytoplankton (Ke, Xie, Guo, Liu, & Yang, 2007). If the grazing pressure from zooplankton on the phytoplankton was reduced by silver carp and bighead carp grazing on zooplankton, there would be a smaller decrease in mean *Microcystis* or total phytoplankton biomass.

Lower pH reduced cyanobacterial growth as lower pH meant there was more highly permeable H_2CO_3 present in the inorganic carbon pool and required higher expenditures of energy by the cyanobacterial cells to maintain enough of it intracellularly so as to carry out photosynthesis ("Cyanobacteria and low pH," n.d.).

Higher water temperature promoted the growth of cyanobacteria, especially *Microcystis* spp. The warmer water allows *Microcystis* to catch up with competitors (Chorus & Welker, 2021). This is because it, as well as many other cyanobacteria, have higher temperature optima than many micro-algae (Chorus & Welker, 2021). Also, "elevated temperatures at the sediment surface may promote the recruitment of cyanobacteria from the sediments" and speed up the degradation of organic matter, leading to more nutrients released for growth (Chorus & Welker, 2021).

Higher precipitation can lead to flushing and de-stratification, which temporarily disrupts cyanobacterial blooms (Reichwaldt & Ghadouani, 2012). By weakening the thermal stratification of the water body, higher precipitation reduced the competitive advantage of buoyant cyanobacteria (Chorus & Welker, 2021).

Higher transparency means that more light is available for the cyanobacteria to do photosynthesis. The fact that higher transparency and thus higher light availability led to a smaller decrease in mean *Microcystis* or total phytoplankton biomass indicated that the phytoplankton community consisted largely of species whose growth rates were highly dependent on light.

Through comparing the regression coefficients, it was clear that variability in pH affected the growth of cyanobacteria the most while the stocking of silver carp at the current biomass was the least significant influencing factor of cyanobacteria. The stocking of bighead carp exerted a greater negative impact on controlling algae bloom than the positive impact exerted by silver carp. Thus, one needs to take careful consideration about the simultaneous use of silver carp and bighead carp and their stocking densities.

4.2 Comparison between Biomanipulation Experiments at Different Lakes

The lower percentage change could be interpreted in two ways: if the percentage change was negative, then the smaller value meant there was more percentage decrease; if the percentage change was positive, then the smaller value meant there was less percentage increase. *Vice versa*.

4.2.1 For Percentage Change in Mean Chlorophyll-a Concentration or Phytoplankton Biomass

Lower transparency (lower light availability) led to less percentage decrease or more percentage increase in cyanobacteria. This might indicate that the cyanobacterial community was dominated by species that were not so dependent on light, and these species were able to better outcompete other phytoplankton organisms at low light intensity.

A larger lake surface area will lead to more pronounced turbulence, particularly if the lake is exposed to wind, and under this condition, even highly buoyant cyanobacteria "have little chance for vertical positioning and are mixed throughout the water column" (Chorus & Welker, 2021). Thus, it would be harder for the cyanobacterial bloom to form. That explained why mean chlorophyll-a concentration or phytoplankton biomass experienced a greater percentage decrease or smaller percentage increase as lake surface area increased.

The depth of a lake is related to its thermal stratification, which refers to the temperature gradients over depth. Deeper lakes develop more stable thermal stratification (Chorus & Welker, 2021). With thermal stratification, warmer water is layered above cooler, denser water (Chorus & Welker, 2021). This provides suitable conditions for cyanobacterial growth as the common bloom-forming types are buoyant and can avoid sedimentation, which allows them to stay in favorable temperatures and accumulate (Chorus & Welker, 2021). Also, more stable stratification leads to reduced mixing (Chorus & Welker, 2021). Under weak turbulence, buoyant cyanobacteria on average might be positioned closer to the surface than their competitors and hence receive more access to light, resulting in better growth (Chorus & Welker, 2021).

Heavier precipitation can reduce thermal stratification so that nutrient-rich water is transported from deeper layers into upper layers where it fertilizes the growth of cyanobacteria (Chorus & Welker, 2021). Hence, more precipitation led to a higher percentage change in mean chlorophyll-a concentration or phytoplankton biomass.

Here, the characteristics of the lake itself, like the depth, or the weather conditions, like precipitation, had a greater impact on how the growth of cyanobacteria varied (reflected by the change in mean chlorophyll-a concentration or phytoplankton biomass) than the stocking of silver carp. Though the treatments with silver carp could decrease cyanobacterial growth to some extent, the current stocking densities were too low to make the influence very significant.

4.2.2 For Percentage Change in Total Nitrogen and Total Phosphorus

The silver carp plays an important role in animal-mediated nutrient recycling. Nitrogen and phosphorus are released back into the lake water through the excretion and egestion of the fish ("Animal-mediated nutrient cycling across lakes | GLEON," n.d.). Therefore, as the silver carp biomass increased, the total nitrogen and total phosphorus decreased less or increased more.

As the depth of the lake increases, thermal stratification becomes more stable. This can reduce mixing and thus reduce the nutrient supply from deeper layers to upper layers (Chorus & Welker, 2021). That explained why in measurement, total phosphorus and total nitrogen decreased more or increased less with the rise in depth.

Heavier precipitation will result in higher flushing. This dilutes out the phosphorus and nitrogen, even causing a loss of nutrients from the enclosures (Chorus & Welker, 2021). Hence, the percentage increase in total phosphorus and

total nitrogen became smaller or the percentage decrease became larger.

There will be more pronounced turbulence in lakes with a larger surface area, particularly if they are exposed to wind (Chorus & Welker, 2021). This meant greater water exchange rates, which made phosphorus and nitrogen less concentrated.

By comparing the absolute value of the regression coefficients, it could be seen that mean lake depth played the most significant role in influencing total phosphorus and total nitrogen. It was also clear that the stocking of silver carp had a greater impact on total phosphorus level than on total nitrogen level. This fits Cui et al. (2004)'s conclusion that “silver carp are more effective at removing phosphorus than nitrogen” (Yi et al., 2016).

4.3 Implications

The regression analysis showed that there was no correlation between cyanobacterial growth and lake TP or TN. This is an unusual finding as total phosphorous concentration is usually used to “estimate the potential of bloom development in a waterbody” and many interventions have been developed to reduce nutrient loads (Chorus & Welker, 2021). The reason behind this result might be that within the trophic states from mesotrophic to hypertrophic, the variation in lake TP and TN does not affect the growth of cyanobacteria. However, to reach a firm conclusion, more experiments at lakes with different trophic levels need to be done.

Since, according to the regression analysis, mean lake depth had the greatest impact on the growth of cyanobacteria, the depth of the lake needs to be taken into consideration when implementing biomanipulation. The shallower the lake, the more effective the biomanipulation will be at controlling cyanobacterial bloom.

Moreover, silver carp is preferred over bighead carp in reducing cyanobacterial bloom. One should be aware that stocking bighead carp will possibly enhance the bloom rather than control it. Also, the silver carp, in general, is less expensive than the bighead carp. The average market price for silver carp is 0.23 US Dollars per pound whereas the average market price for bighead carp is 0.48 US Dollars per pound (Towers, 2010; Stone, Engle, Heikes, & Freeman, 2000). Therefore, the cost for biomanipulation will be lower using silver carp only rather than using a combination of silver carp and bighead carp. However, even if only silver carp is used, the current biomass is not sufficient to control cyanobacterial blooms effectively. Nevertheless, one should be also aware that the silver carp is considered an invasive species in the Great Lakes of the United States, competing with many native fish species for food ("Carp -Silver," n.d.). Therefore, the implementation of biomanipulation requires a comprehensive understanding of the lake's food web to avoid disruption to the local aquatic ecosystem.

Through the comparison between experiments within lakes, it was found that change in pH had the greatest impact on the growth of cyanobacteria. There was a positive feedback loop where higher pH enhanced cyanobacterial growth and cyanobacterial bloom increased pH through the uptake of hydrogen carbonate during intense photosynthesis, which shifted the equilibrium between carbonate and hydrogen carbonate and made the water even more alkalic (Chorus & Welker, 2021). However, one should be very cautious about pH adjustment as an intervention to control cyanobacterial growth in natural water bodies like freshwater lakes. Chemical manipulation of lakes may bring about unintended

consequences, though these consequences still remain unclear. Further research and experiments can be done to investigate the effects of chemical manipulation interventions.

Additionally, it is important to consider the influence of climate change on cyanobacterial blooms. Our analysis showed that for Lake Taihu, a higher water temperature increased cyanobacterial growth, while across different lakes, the water temperature had no effect on the blooms. Precipitation also acted as an influencing factor, either enhancing or reducing cyanobacterial proliferation depending on the specific conditions at the lakes. Therefore, how climate change will affect cyanobacteria is far more complex than the direct impact of certain weather parameters, and further investigations can be done to gain a more thorough understanding of climate change's effects on cyanobacterial blooms in both the present and the future.

Admittedly, this paper only compared four different lakes, which was a relatively small sample size. To get a more well-rounded view of what factors influence cyanobacterial growth and the effectiveness of non-classic biomanipulation, more lakes need to be studied. Moreover, wind also has an impact on the growth of cyanobacteria and the formation of algae bloom. However, to study the effect of wind, wind direction, which is changing all the time, needs to be taken into consideration. This will bring too much variability into the analysis. A more focused study on the effect of wind speed and wind direction on cyanobacteria can be done in the future.

References

Akyuz, D., Luo, L., & Hamilton, D. (2014). Temporal and spatial trends in water quality of Lake Taihu, China: analysis from a north to mid-lake transect, 1991–2011. *Environmental Monitoring And Assessment, 186*(6), 3891-3904. doi:10.1007/s10661-014-3666-0

Animal-mediated nutrient cycling across lakes | GLEON. (n.d.). Retrieved June 26,2022, from https://gleon.org/research/projects/animal-mediated-nutrient-cycling-across-lakes

Bartram, J. (Eds.). (2015). *Routledge Handbook of Water and Health*. London: Routledge

Biomanipulation. (2005). Retrieved July 22, 2022, from https://www.sciencedirect.com/topics/earth-and-planetary-sciences/bio-manipulation#:%7E:text=Biomanipulation%2C%20a%20term%0introduced%20by%20Shapiro%20et%20al.,to%2C%20or%20to%20suppla%2C%20reductions%20of%20nutrient%20loading

Carp - Silver. (n.d.). Retrieved July 22, 2022, from https://www.michigan.gov/invasives/id-report/fish/carp-silver

Chorus, I., & Welker, M. (Eds.). (2021). *Toxic Cyanobacteria in Water* (2nd ed.). CRC Press.

Cui, F. Y., Lin, T., Ma, F., & Zhang, L. Q. (2004). Experimental studies on biomanipulation of silver carp and bighead carp in water resources management. *J.Nannjing Univ. Sci. Technol, 28*, 668-692.

Cyanobacteria and low pH. (n.d.). Retrieved July 20, 2022, from https://www.researchgate.net/post/Cyanobacteria_and_low_pH

eutrophication | Definition, Types, Causes, & Effects. (n.d.). Retrieved July 17, 2022, from https://www.britannica.com/science/eutrophication

Eutrophication. (n.d.). Retrieved July 17, 2022, from https://www.coursehero.com/file/75050126/ttppprtf/

Guo, L., Wang, Q., Xie, P., Tao, M., Zhang, J., Niu, Y., & Ma, Z. (2014). A non-classical biomanipulation experiment in Gonghu Bay of Lake Taihu: control of Microcystis blooms using silver and bighead carp. *Aquaculture Research,46*(9),2211-2224. doi: 10.1111/are.12375

Lake Donghu | Donghu | World Lake Database - ILEC. (n.d.). Retrieved June 26, 2022, from https://wldb.ilec.or.jp/Display/html/3529

Ke, Z., Xie, P., Guo, L., Liu, Y., & Yang, H. (2007). In situ study on the control of toxic *Microcystis* blooms using phytoplanktivorous fish in the subtropical Lake Taihu of China: A large fish pen experiment. *Aquaculture, 265*(1-4), 127-138. doi:10.1016/j.aquaculture.2007.01.049

Qin, B., Paerl, H., Brookes, J., Liu, J., Jeppesen, E., & Zhu, G. et al. (2019). Why Lake Taihu continues to be plagued with cyanobacterial blooms through 10 years (2007–2017) efforts. *Science Bulletin, 64*(6), 354-356. doi:10.1016/j.scib.2019.02.008

Reichwaldt, E., & Ghadouani, A. (2012). Effects of rainfall patterns on toxic cyanobacterial blooms in a changing climate: Between simplistic scenarios and complex dynamics. *Water Research, 46*(5), 1372-1393. doi: 10.1016/j.watres.2011.11.052

Stone, N., Engle, C., Heikes, D., & Freeman, D. (2000, September). Bighead Carp. Retrieved July 22, 2022, from https://wkrec.ca.uky.edu/files/bigheadcarp.pdf

Tang, H., Xie, P., Lu, M., Xie, L., & Wang, J. (2002). Studies on the Effects of Silver Carp (*Hypophthalmichthys molitrix*) on the Phytoplankton in a Shallow Hypereutrophic Lake through an Enclosure Experiment. *International Review Of Hydrobiology, 87*(1), 107. doi: 10.1002/1522-2632(200201): 87:1<107::aid-iroh107> 3.0.co;2-j

Towers, L. (2010, March 2). How to farm silver carp. Retrieved July 22, 2022, from https://thefishsite.com/articles/cultured-aquatic-species-silver-carp

Vollenweider, R. A., & Kerekes, J. (1982). *Eutrophication of waters: Monitoring, assessment and control*. Paris: Organization for Economic Co-operation and Development.

Ye, R., Shan, K., Gao, H., Zhang, R., Wang, S., & Qian, X. (2015). Long-term seasonal nutrient limiting patterns at Meiliang Bay in a large, shallow and subtropical Lake Taihu, China. *Journal Of Limnology*, (AoP). doi: 10.4081/jlimnol.2015.1147

Yi, C., Guo, L., Ni, L., Yin, C., Wan, J., & Yuan, C. (2016). Biomanipulation in mesocosms using silver carp in two Chinese lakes with distinct trophic states. *Aquaculture, 452*, 233-238. doi: 10.1016/j.aquaculture.2015.11.002

Yin, C., Guo, L., Yi, C., Luo, C., & Ni, L. (2017). Physicochemical Process,Crustacean, and *Microcystis* Biomass Changes in situ Enclosure after Introduction of Silver Carp at Meiliang Bay, Lake Taihu. *Scientifica, 2017*, 1-9. doi:10.1155/2017/9643234

Ye, R., Shan, K., Gao, H., Zhang, R., Wang, S., & Qian, X. (2015). Long-term seasonal nutrient limiting patterns at Meiliang Bay in a large, shallow and subtropical Lake Taihu, China. *Journal Of Limnology*, (AoP). doi:10.4081/ jlimnol. 2015.1147

Zhang, X., Xie, P., Hao, L., Guo, N., Gong, Y., Hu, X., . . . Liang, G. (2006). Effects of the phytoplanktivorous Silver Carp (*Hypophthalmichthys Molitrixon*) on plankton and the hepatotoxic microcystins in an enclosure experiment in a eutrophic lake, Lake Shichahai in Beijing. *Aquaculture, 257*(1-4), 173-186. doi: 10.1016/j.aquaculture. 2006.03.018

Dammed Rivers: Hydropower Development Modifies Fish Community Dynamics in the 3S Basin of Mekong

Yanzhi Chen

Author Background: *Yanzhi Chen grew up in China and currently attends Keystone Academy in Beijing, China. Her Pioneer research concentration was in the field of environmental studies/ecology and titled "Waste or Not Waste, that's the Question – Solving Major Environmental Problems."*

Abstract

The Mekong, one of the world's most biodiverse rivers, has undergone rapid hydropower construction in the past decades. By disconnecting tributaries, disrupting flow regimes, changing sediment load, and deteriorating water quality, hydropower stations drastically affect hydrology and subsequently the fish community. Focusing on one of the richest and most rapidly developing Mekong subregions, the 3S basin (including the Sesan, Sre Pok, and Sekong tributaries), the study examined how dam construction alters hydrological dynamics and thus the interactions between hydrology and fish communities. The study found strong evidence that major dam construction decreases fish abundance and richness and shifts hydrological characteristics in flow, sediment load, water quality, and water temperature. The study also found that fish richness and abundance correlate with hydrological characteristics both concurrently and in a time-lagged manner, while dam construction modifies this relationship between fish and hydrology, potentially causing far-reaching consequences. The results strongly call for a reassessment of dams planned in the 3S basin.

1. Introduction

As the world's second-most biodiverse river (Ziv *et al.*, 2012) and one of the 35 biodiversity hotspots (Mittermeier *et al.*, 2011), the Mekong River system is one of the most important freshwater aquatic ecosystems in the world (Soukhaphon *et al.*, 2021). Unparalleled in its biodiversity and ecosystem services, the Mekong hosts 894 indigenous fish species in 24 orders and 87 families (Foese & Pauly, 2010) with its higher taxonomic diversity paramount among all rivers (Valbo-Jørgensen *et al.*, 2009). New species are still discovered at a rapid rate (Signs *et al.*, 2022), marking the system as a region with highly understudied fish communities.

However, the Mekong River system is threatened by 89 dams in operation, 28 dams under construction, and 102 dams being planned that are larger than 15MW

(Fig. 1) (Winemiller *et al.*, 2016; Eyler, 2020). The total active reservoir storage is projected to increase from occupying just 2% of the annual discharge in 2008 to 19% in 2025 (Kummu & Varis, 2007; Hecht *et al.*, 2018), but the cascading repercussions to biodiversity induced by degraded ecosystems and fragmented rivers are often overlooked (Winemiller *et al.*, 2016). Rapid hydropower development changes abiotic components of the river, such as flow regime (e.g., Lu & Chua, 2021), sediment discharge (e.g., Kondolf *et al.*, 2014), water temperature (e.g., Bonnema *et al.*, 2020), and water quality (e.g., Sor *et al.*, 2021), altering habitats and affecting fish abundance, diversity, and distribution (Li *et al.*, 2013; Ngor *et al.*, 2018).

Fish diversity in the Mekong is characterized by large-scale seasonal migrations (Poulsen *et al.*, 2004), which can be severely affected by decreased river connectivity due to dam construction (Ziv *et al.*, 2012; Ou & Winemiller, 2016). In the Mekong region, river connectivity decreased from 93% to 77% from 1990 to 2010 and is projected to drop to 10% in worst-case scenarios by 2022 (Grill *et al.*, 2014). River fragmentation blocks important headwater spawning grounds (Hogan *et al.*, 2007; Altermatt & Fronhofer, 2018), isolates fish communities to change species composition (Ganassin *et al.*, 2021), and obstructs the formation of refuge habitats and species pools (Shao *et al.*, 2019). Such effects may be intensified by the fact that many dams are built near headwaters (Grill *et al.*, 2014), which often serve as spawning sites (MRC, 2006). However, apart from these direct influences, which are studied relatively comprehensively in the Mekong basin (Ziv *et al.*, 2012), dam construction in the region causes hydrological changes in flow, sediment load, and water quality, which indirectly reduces fish abundance and diversity.

As a river system characterized by regular and distinct interannual flow patterns (Ou & Winemiller, 2016), 75% of the annual discharge passes through the delta during the wet season from June to November, which is pivotal for sustaining the rich ecosystem (Hecht *et al.*, 2018). Hydropower dam constructions are projected to disrupt the frequency, magnitude, timing, duration, and rate of change of the flow patterns (Poff & Ward, 1989; Huang *et al.*, 2011), subsequently altering fish habitats by disrupting the ecological integrity of the aquatic system (Poff *et al.*, 1997). Seasonal flow in the LMB can both be attenuated or weakened depending on the location and type of hydropower installation (Hecht *et al.*, 2018; Lu & Chua, 2021). The natural streamflow determines aspects such as habitat diversity, water temperature, water quality, and connectivity, thus playing important roles in determining ecosystem dynamics (Poff *et al.*, 1997; Wolman and Miller, 1960). The life cycles of numerous freshwater species depend on diverse spatial-temporal habitats, which can be provided by the variable hydraulics (Sparks, 1995). Additionally, regular flow patterns foster certain riparian vegetation, which provides nutrients and special habitats for the inland aquatic lives (Nilsson *et al.*, 1991).

By trapping sediments in the reservoirs, dams influence the total suspended solid concentration and thus sediment deposition in the floodplain, which leads to sediment starvation (Kondolf *et al.*, 2014). Models and empirical data support sediment loss due to hydropower construction (Xue *et al.*, 2011; Kondolf *et al.*, 2014; Mahn *et al.*, 2015), and sedimentation is predicted to be lowered by 90-96% in the delta in the next decades (Mahn *et al.*, 2015; Kondolf *et al.*, 2014). Sediments carry essential nutrients, which foster the growth of micro-and macro-organisms that are important for fish productivity (Kummu & Varis, 2007). Additionally, the traditionally high sediment load of 160 Mt per year in the Mekong Delta provides a turbid habitat to which most fishes have adapted (Valbo-Jørgensen *et al.*, 2009). Rapid changes in sediment load, therefore, are projected to influence fish abundance and diversity.

Hydropower dam construction also triggers changes in water quality, i.e., physicochemical aspects of dissolved oxygen, heavy metal concentration, and pH (Fantin-Cruz *et al.*, 2015). Upstream of the dam, impoundment transforms the reservoir from a lotic to a lentic habitat, which decreases flow velocity, degrading the river's self-purification functions and increasing nutrient and heavy metal concentrations (Fan *et al.*, 2015; Wei *et al.*, 2009). Below the dam blockage, the combined effects of the hydropower regime and sediment entrapment by the hydropower stations decrease sediment flow, thus reducing nutrients and other heavy metals (Trung *et al.*, 2018; Sor *et al.*, 2021) and causing salt intrusion in the Mekong delta (Eslami *et al.*, 2019). Fish are sensitive to environmental quality changes (Basavaraja *et al.*, 2014). Low levels of dissolved oxygen reduce fish survival whereas excessive concentrations of ammonia and phosphorous may lead to eutrophication (Schindler, 1974). Additionally, phytoplankton and zooplankton respond rapidly to water quality changes (Webber *et al.*, 2005), which also influence primary productivity and thus fish productivity in the river (Sor *et al.*, 2020; Ngor *et al.*, 2018).

Despite the importance of this topic, most papers focus on a one-way influence of hydrological factors on fish, or dams on hydrology, whereas few studies address the three-way temporal connection between fish biodiversity, dam construction, and hydrological regimes. Some research papers investigate the impacts of flow regimes on fish abundance, diversity, and assemblages (e.g., Ngor, 2018), but other hydrological factors are not assessed systematically, and therefore, there is a lack of holistic assessments of how multiple factors collectively shape fish productivity and diversity in the Mekong basin.

This study covers three major Mekong tributaries in the 3S Basin: The Sesan, Sre Pok, and Sekong rivers (Fig. 2). The 3S system is both a hotspot for fish biodiversity and dam construction (major dams shown in Fig. 2), characterized by high fish endemicity (Winemiller *et al.*, 2016) and high migratory fish counts (IUCN, 2013). Meanwhile, the unprecedented hydropower developments in the basin (Null *et al.*, 2020; Baran *et al.*, 2015) are projected to severely affect the fish populations (Piman *et al.*, 2016; Ziv et al., 2012). Studies found hydrological changes, including a 98% increase in dry-season flow (Piman *et al.*, 2016), decreased flow duration (Chantha & Ty, 2020), a 40-80% reduction in annual suspended sediment load (Wild & Loucks, 2014), and degradation in water quality (Sor *et al.*, 2021), which may explain the decrease in fish abundance and diversity in the basin (MRC, 2021; Ngor *et al.*, 2016). Thus, an investigation of the 3S system is timely and needed.

In this paper, I explore how dam construction alters the patterns between hydrological factors and fish abundance and diversity in the 3S system. I will test three hypotheses: 1) major dam construction shifts the level and seasonal dynamics of hydrology and fish richness and abundance in the 3S system, 2) fish abundance and diversity are closely correlated to lagged and immediate changes in hydrology, and 3) dam construction alters the dynamics between hydrology and fish communities, which vary across rivers.

2. Study Area

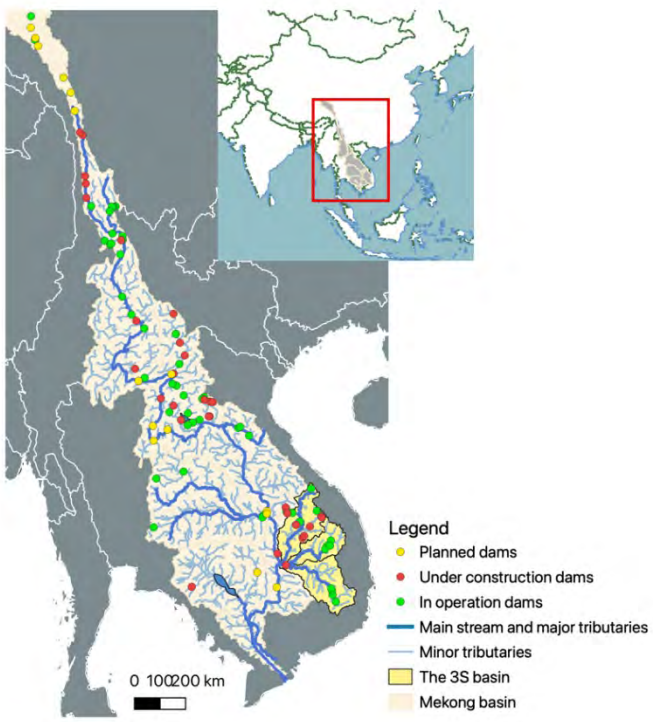


Figure 1. *Planned, in operation, and under construction dams in the Mekong Basin as of 2020.*

The Mekong extends from Southwestern China, across Thailand, Lao PDR, Cambodia, and Vietnam towards the South China Sea, stretching 4,000 kilometers (Ziv *et al.*, 2012). As a tropical floodplain river (Hecht *et al.*, 2018), it is characterized by definite wet and dry monsoonal seasons, substantial sediment discharge, and highly migratory fish populations (Montana *et al.*, 2020).

The 3S, situated in the Lower Mekong Basin (Fig. 2), covers 78,650 square kilometers (MacQuarrie *et al.*, 2013). All three tributaries originate in the Vietnam highlands and merge with the main river at Stung Treng in Cambodia. The Sesan originates in Vietnam and joins the Sekong 15 kilometers before Stung Treng, while the Sre Pok joins with the Sesan 30 kilometers before Stung Treng (Bunthang & Phen, 2021). Of the 149 species found in the 3S basin, 42% belong to the highly migratory *Cyprinidae* family, followed by *Balitoridae* (17%), and *Cobitidae* (11%) (Valbo-Jørgensen *et al.*, 2009). Migratory fish display a gradient of richness from the higher altitudes in Vietnam to Cambodia (Constable, 2015; Chea *et al.*, 2016a) (Fig. 2). Figures are created using QGIS (v. 3.0, QGIS Development Team, 2022).

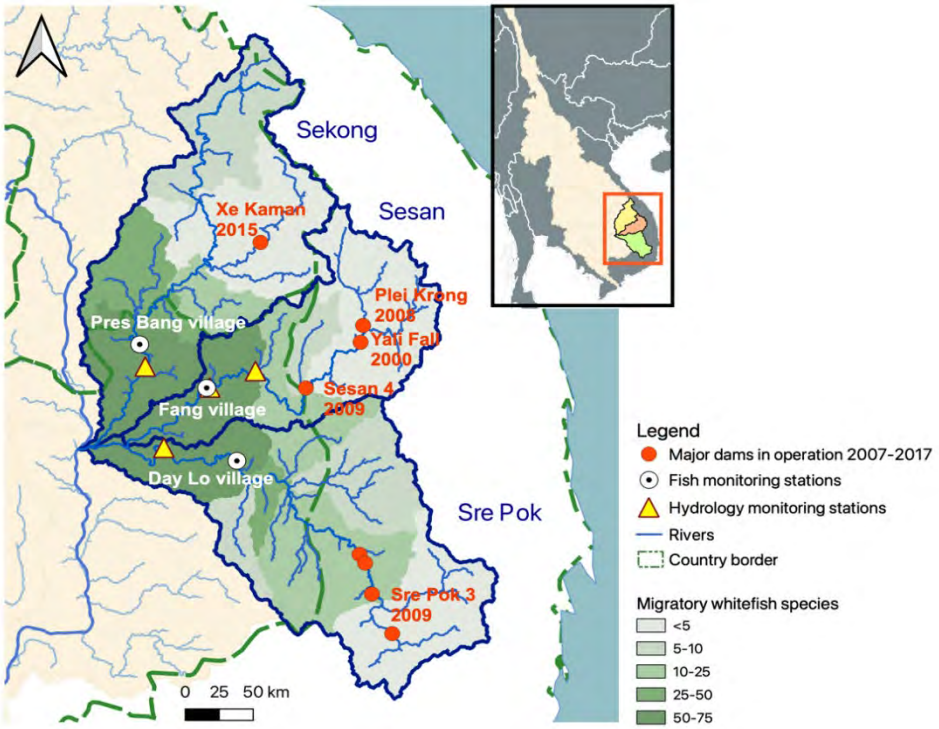


Figure 2. *Monitoring stations for fish and hydrology, along with the major dams in operation from 2007 to 2017 in the 3S basin. Gradient shows the richness of migratory fish in the basin.*

During the study period from 2007 to 2017, four major dams (following the categorization by Bonnema *et al.*, 2015) and numerous minor dams were constructed (Table 1).

Table 1. *Hydropower dams in operation during the study period of 2007-2017, listing the main characteristics of the location, commission year, installed capacity, active storage, and reservoir area (modified from Piman et al., 2016; Räsänen et al., 2014; Bonnema et al., 2020).*

Hydropower Project	River basin	Commission Year	Installed Capacity (MW)	Active Storage (Mm^3)	Catchment Area (km^3)
Plei Krong	Sesan	2008	100.0	948.0	3216.0
Sesan 4	Sesan	2009	360.0	264.2	9326.0
Xe Kaman	Sekong	2015	290.0	1683.0	3580.0
Sre Pok 3	Sre Pok	2009	220.0	62.6	9410.0

3. Method

3.1. Data collection

Hydrology and fish catchment data are monitored from 2007 to 2017. Fish abundance and richness are calculated from the recorded fish caught at the Day Lo village in Lum Phat, Ratanakiri (Sre Pok), Pres Bang village in Siem Pang, Stung Treng (Sekong), and Fang village in Veounsi, Ratanakiri (Sesan). Hydrological characteristics are measured at Siempang station and Stung Treng station (Sekong), Lumphat station (Sre Pok), Andaung Meas station, and Voeun Sai station (Sesan) (Fig. 2).

Fish catchment data collection follows the standard Mekong River Commission (MRC) procedures (MRC, 2007). Stationary gillnets (length: 120 $\pm$ 50 m; height: 2–3.5 m; mesh size: 3–12 cm; soak h/d: 12 $\pm$ 2) are deployed by three fishers at each site every day. The fishermen also employ other techniques to catch fish, but only stationary gillnet data is used for consistency. The species are identified by the trained fishers using the MRC fisheries database, supervised by the fishery researchers at the Inland Fisheries Research and Development Institute of the Cambodia Fisheries Administration (Nuon *et al.*, 2020). Data are recorded daily on the species, count, fish weight, and the number of hours fishing.

The water level is recorded in-situ by direct measurements of water height in the Discharge-Sediment Monitoring Project twice per day (MRC, 2014). According to the standard MRC procedure (MRC, 2017), water quality parameters of water temperature (T), pH, total suspended solids (TSS), dissolved oxygen (DO), total phosphorous (T-P), ammonium (NH_4-N), and total nitrate-nitrite concentration ($NO_{2\text{-}3}$-N) are measured once a month using a surface grab that collects water samples at 30 to 50 centimeters below the surface in the middle of the free-flowing stream (MRC, 2017). For the latter five parameters where in-situ measurements are not possible, samples are taken with proper preservative agents and stored at a cool temperature in labs for further analysis.

3.2. Data analysis

Fish data are assessed with richness and abundance measures. Fish richness is calculated as the aggregate number of species caught per week using the specnumber function in the package *vegan* in R (Oksanen *et al.*, 2022). Fish abundance data is calculated as the catchment per unit effort (CPUE), or the average number of fish divided by the number of fishing hours for each net. Before analysis, the Inverse Simpsons Index is calculated for each river to provide a sense of fish richness and equality using the *vegan* package (Oksanen *et al.*, 2022).

To keep a consistent temporal grain, all parameters are averaged to monthly values.

3.3. Modifications of hydrological parameters and fish community by dam construction

To understand how dam construction alters hydrology and fish community in the 3S system along the temporal axis, violin plots from the package *vioplot* are used to compare each parameter before and after a major dam establishment event at each site (Adler & Kelly, 202). The violin plot depicts a density distribution curve of the series while highlighting the 25th, 50th, and 75th quantile values. Each river is split into

a pre-dam and a post-dam period by the construction date of the major dams listed in Table 1. For the Sesan river, with two dams being constructed in 2008 and 2009, the split is set at the construction date of the latter in early 2009. To standardize the axis for comparison, all parameter data are standardized to a mean of 0 and a standard deviation of 1. The non-standardized median values before and after the major dam construction and percent changes are also calculated to contextualize the degree of change. Results are cross-verified with time-series graphs showing the temporal dynamics of each parameter created using *ggplot2* (Wickham, 2016).

Using ggseasonplot in *ggplot2*, variables showing moderate to strong seasonality are identified (Wickham, 2016) and boxplots by month are plotted showing the distribution of the values before and after dam construction across the year, to understand how dam construction shifts the seasonal dynamics.

3.4. Correlation between hydrological factors and fish

Multiple linear regression maps a linear combination of variables that explains the response variables with least-squares fit (Tranmer *et al.*, 2020) using the equation,

$$Y_1 = b_0 + b_1X_1 + b_2X_2 + \cdots + b_kX_k + \varepsilon_i$$

Where Y_1 is the response variable, b_0 is the intercept term, b_i is the slope coefficient for the explanatory variable X_i, and ε_i is the error term. The regression models how hydrology correlates to either fish abundance or diversity for each river. Explanatory variables are first plotted, and variables with high autocorrelation ($R \geq 0.7$) are excluded from the regression analysis. The hydrological factors with the highest *p-value* in each model are excluded successively until the combination of parameters that yields the lowest overall *p-value* is reached. This model is selected and reported as the best-performing model. The null hypothesis, H_0, which assumes no correlation between the hydrological parameters and fish abundance or richness, is rejected when the $p\text{-}value \leq 0.05$. The normal distribution of residuals and homoscedasticity of the regression are tested using the P-P plot and the plot of standardized residuals against the standardized predicted values, respectively.

3.5. Lagged time series regression

A cross-correlation function (CCF) is a cross-product correlation calculated as a function of lag between two series (Bennett, 1979). As changes in hydrology may take effect in fish communities after months (Pyron *et al.*, 2006; Fornaroli *et al.*, 2020), the CCF tests for the time-lagged effect of the hydrological factors on fish communities and determines the best predicting lag of the explanatory variable. The lag2.plot function from package *astsa* is applied between each hydrological parameter and either fish abundance or richness, creating a grid of scatterplots lagging from 0 to 12 months (Stoffer & Poison, 2022). A LOWESS smoothing line is also plotted in each grid to identify the direction of the correlation. The optimal lag is determined with the maximum correlation coefficient, R.

Multiple regression with lag measures how the response variable, richness and abundance, can be predicted by leading *x*-variables with different time lags, represented by the equation,

$$Y_1 = b_0 + b_1X_1\left(t - t'_{X_1}\right) + b_2X_2\left(t - t'_{X_2}\right) + \cdots + b_iX_i\left(t - t'_{X_i}\right) + \varepsilon_i$$

where t'_{X_i} is the lag time for the variable X_i.

From the previous CCF analysis, variables with a significant correlation coefficient at the optimal lag ($R \geq 0.2$) are used in this lagged regression. Similarly, the model of a combination of lagged variables with the lowest *p-value* is selected and reported.

3.6. Lagged time series regression before and after dam construction

For Sre Pok and Sekong, where considerable data before and after dam construction can be obtained, lagged multiple regressions are conducted separately for data before and after dam construction to investigate the differences in the response of fish richness and abundance to hydrological factors due to dam construction. After identifying significant lag variables using the CCF function for each dataset, multiple regression is conducted and the combination of leading variables with the lowest *p-value* is reported.

All calculations are conducted in R (v. 4.0, R Core Development Team, 2021).

4. Results

All data were collected from June 1st, 2014, to December 31st, 2017. The data included 16,640, 36,948, and 17,268 recordings of fish caught by type per day at Pres Bang (Sekong), Day Lo (Sre Pok), and Fang (Sesan) villages, with the Inverse Simpson's Index of 31.562, 17.985, and 17.857, respectively. For the explanatory variables, 7992 recordings of water level (WL), 110 recordings of total suspended solids (TSS), 113 recordings of total phosphorous (T-P), 113 recordings of dissolved oxygen (DO), 113 recordings of nitrite-nitrate ($NO_{2\text{-}3}$-N), 113 recordings of total ammonium (NH_4-N), 113 recordings of water temperature (T), and 101 recordings of pH are obtained at each river. Weekly fish richness ranges from 8 to 87 species, and fish abundance ranges from 0.22 to 13.93 fish per hour per net. WL range from 1.53 to 9.89 m, TSS from 2 to 350 mg/L, T-P from 0.001 to 0.538 mg/L, DO from 3.99 to 10.91, $NO_{2\text{-}3}$-N 0.001 to 0.596 mg/L, NH_4-N from 0 to 0.484 mg/L, T from 23.5 to 38℃, and pH from 5.31 to 8.39.

In the CCF analysis, non-consecutive missing data between months are smoothed by taking the mean between the previous and the next recording. The weighted average across more values is not taken due to the relatively large degree of fluctuation between each measurement. The data for total suspended solid concentration is disregarded in all time-series related analyses due to too many (9 out of 12 months) missing data in 2009, which makes smoothing impossible. Since water level data shows distinct seasonality (Supplement 1, Fig. 3, 13, 23), its consecutive missing data are smoothed by taking the average of the recording at the same month in the previous and the next year.

The Lower Sesan 2 dam, the largest dam in downstream Sesan, started construction in 2014 (Ziv *et al.*, 2012), which turned the river into a large reservoir, significantly altering hydrological patterns and fish communities (Baird, 2014). Patterns of fish abundance and richness show abnormity (S1, Fig. 5 and 6), and the high disturbance also makes the comparison between fish dynamics before and after the completed dams inaccurate. Therefore, this paper examines the impacts of hydrology at Sesan from 2007 to 2014, and the same set of analyses for the whole study period at Sesan is included in S3 for reference.

4.1. Modifications of hydrological parameters and fish community by dam construction

Figure 3 shows the changes in fish and hydrology before and after dam construction, and Table 2 shows the changes in the median values.

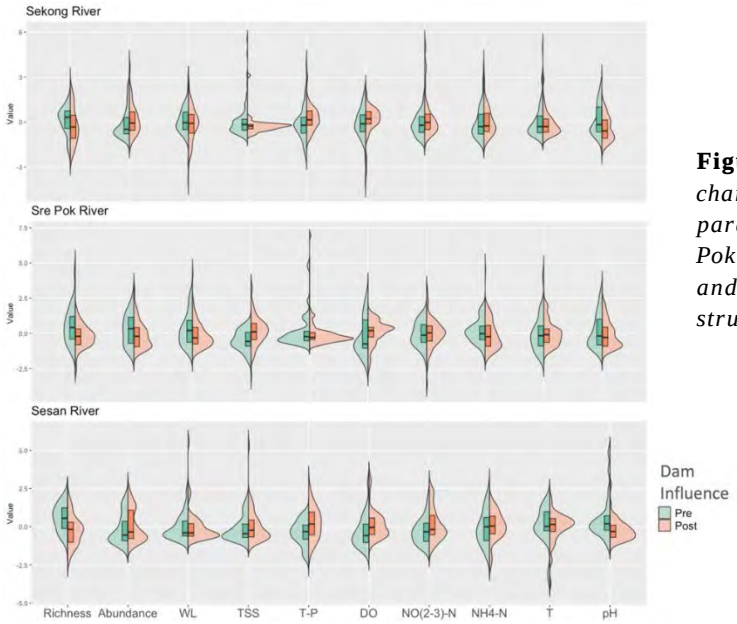


Figure 3. *Violin plots showing the changes in fish and hydrological parameters at Sekong (3a), Sre Pok (3b), and Sesan (3c) before and after the major dam construction event.*

Table 2. *Median values and percent changes in the hydrological factors and fish indices before and after dam construction. Significant changes (>10%) are bolded and negative changes are highlighted in red. Values are rounded to four significant figures and percent changes are rounded to two decimal places.*

	Sekong			Sre Pok			Sesan		
Variable	**Before**	**After**	**Percent Change (%)**	**Before**	**After**	**Percent Change (%)**	**Before**	**After**	**Percent Change (%)**
Richness	**46.0**	**36.0**	**-21.7**	**50.0**	**44.0**	**-12.0**	**33.0**	**27.5**	**-16.7**
Abundance	**1.11**	**1.25**	**12.6**	**1.12**	**.995**	**-10.8**	**3.87**	**2.76**	**-28.7**
WL (m)	**3.54**	**3.08**	**-12.9**	3.43	3.23	-5.8	**3.32**	**3.75**	**13.2**
TSS (mg/L)	41.0	41.0	0	**52.0**	**34.8**	**-33.1**	22.0	22.0	0
T-P (mg/L)	**.0606**	**.0948**	**56.3**	**.125**	**.0799**	**-36.0**	**.046**	**.065**	**41.8**
DO (mg/L)	7.52	7.47	0.6	7.166	7.834	9.3	7.29	7.69	5.5
NO2-3-N (mg/L)	.0787	.0820	4.3	.153	.161	5.4	**.0909**	**.131**	**44.0**
NH4-N (mg/L)	**.0440**	**.0378**	**-14.0**	.0365	.0328	-10.0	**.0353**	**.0407**	**15.4**
T (degC)	28.7	29.5	2.8	27.0	28.6	5.9	29.0	29.1	0.3
pH	7.08	7.20	1.7	7.05	7.12	1.0	7.00	7.06	0.9

Among all rivers, richness decreased after dam construction. This can be most clearly observed at the Sekong river, where the time series shows a sudden decrease in fish richness in the year following dam construction (S1, Fig. 1). At two of the three rivers (Sre Pok and Sesan), fish abundance by CPUE also decreased. Changes in hydrological factors are more variable. This is evident as T-P decreases at Sre Pok but increases at Sekong and Sesan, $NO_{2\text{-}3}$-N concentration increased significantly at Sesan and weakly at the other two rivers, and NH_4-N increased at Sesan and decreased at Sekong. TSS decreases at Sre Pok, DO weakly increases at Sre Pok and Sesan, and pH increases at Sekong, while T increases at all sites.

The WL displays a consistent seasonal pattern with peaks in the wet season from July to September and troughs in the dry season from February to April. The fluctuations of TSS and T display a weak seasonal pattern, with TSS fluctuating similarly to WL, and T showing an inverse pattern (S1, Fig. 31-33). Analysis of seasonal dynamic changes shows that at all three rivers, the WL during the wet season decreases. The magnitude of fluctuation in WL is reduced at Sekong and Sre Pok as WL in the wet season decreases and levels in the dry season remain the same or increase (Fig. 4a). Additionally, at the Sesan river, there are more outliers during the post-dam period towards higher WL, which may indicate disruption of the regular flow pattern by dam construction. Seasonal patterns of TSS at the three rivers are consistent before and after dam construction, but its concentration is higher at Sre Pok with a notable increase in May pre-dam construction (Fig. 4b). The fluctuations in T decreased post-dam construction at Sekong and Sre Pok. Notably, at Sre Pok, T displays minimal fluctuations post-dam while there were distinct seasonal patterns before dam construction (Fig. 4c). The rest of the figures are shown in S2.

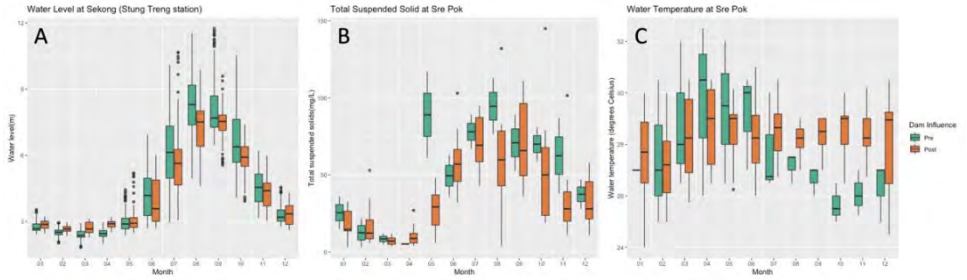


Figure 4. *The changes in the seasonal dynamics of selected parameters at Sekong, Sre Pok, and Sesan. 4a) the change in water level at Sekong before and after dam construction in 2015. 4b) The seasonal change in total suspended solid concentration at Sre Pok before and after dam construction. 4c) The seasonal change in water temperature at Sre Pok before and after dam construction.*

4.2. Correlation between hydrological factors and fish

None of the hydrological parameter pairs have a correlation coefficient higher than 0.8, thus all values are kept in all regression models. Fish abundance data at the three sites are log-transformed in all regressions as they do not fit the assumptions of normality. With the transformed data, all multiple regressions fit the assumptions of a regression analysis tested using the normal Q-Q plots and the plot of standardized residuals against the standardized predicted values.

In the regression model without lags, fish richness at Sre Pok and fish abundance at Sre Pok and Sekong show a statistically significant correlation to the hydrological parameters of WL, DO, and NO_{2-3}-N for fish richness, and WL, TSS, NO_{2-3}-N, and pH for fish abundance (Table 3). Generally, WL and NO_{2-3}-N are strong governing factors of fish richness and abundance as they occur in four out of the six optimized regression models. NO_{2-3}-N appears to be a strong control for fish richness at all sites, which mostly correlates negatively to fish richness. DO and pH appear in two out of the six models yet cause inverse changes to fish richness and abundance at different sites. The full models are shown in S2.

4.3. Lagged time series regression

At the Sekong river, time-lagged pH negatively correlates to fish richness; lagged WL negatively correlates to fish abundance and the lagged T-P positively correlates to fish abundance (Table 4). At Sre Pok, lagged WL and non-lagged DO negatively correlate to fish richness, while lagged NO_{2-3}-N positively correlates. Conversely, lagged WL and lagged NO_{2-3}-N correlate negatively to fish abundance while lagged T-P and lagged pH correlate positively to fish abundance. At Sesan from 2007 to 2014, time-lagged NH_4-N and lagged temperature negatively correlates to fish richness, while lagged NO_{2-3}-N and non-lagged NH_4-N correlate negatively to fish abundance.

Compared to the baseline models, almost all the *p-values* decreased, indicating closer fits. All models except fish richness at Sekong are statistically significant.

Table 3. *Multiple regression describing the influence of hydrology to fish community, with selected variables to minimize the p-value. Significant results are bolded. Significant figures are kepted according to Simpsons et al., (1960) and Agu*inis *et al.* (2021).

	Fish Richness (Sekong)		**Fish Abundance (Sekong)**		**Fish Richness (Sre Pok)**	
Variable	Coefficient (Standard Error)	*p-value*	Coefficient (Standard Error)	*p-value*	Coefficient (Standard Error)	*p-value*
WL	-0.04 (0.111)	0.069	-0.35 (0.136)	0.012	0.18 (0.096)	0.069
TSS	--	--	0.21 (0.133)	0.123	--	--
T-P	--	--	--	--	--	--
DO	0.15 (0.221)	0.084	--	--	-0.33 (0.099)	0.002
NO_{2-3}-N	0.22 (0.127)	0.060	--	--	-0.18 (0.096)	0.060
NH_4-N	--	--	--	--	--	--
T	--	--	--	--	--	--
pH	-0.21 (0.110)	0.061	--	--	--	--
Adjusted R^2	0.064		**0.051**		**0.160**	
p-value	0.053		**0.040**		**0.001**	

Variable	Fish Abundance (Sre Pok)		Fish Richness (Sesan)		Fish Abundance (Sesan)	
	Coefficient (Standard Error)	*p-value*	Coefficient (Standard Error)	*p-value*	Coefficient (Standard Error)	*p-value*
WL	--	--	-0.16 (0.077)	0.042	--	--
TSS	--	--	--	--	--	--
T-P	--	--	0.13 (0.073)	0.081	--	--
DO	--	--	--	--	--	--
NO_{2-3}-N	-0.20 (0.107)	0.069	-0.12 (0.067)	0.077	--	--
NH_4-N	--	--	--	--	-0.28 (0.123)	0.029
T	--	--	--	--	--	--
pH	0.138 (0.108)	0.204	--	--	--	--
Adjusted R^2	**0.053**		0.090		0.075	
p-value	**0.046**		0.060		0.030	

Table 4: *Multiple regression describing the influence of time-lagged hydrology to fish community, with selected variables to maximize the p-value. Significant results are bolded.*

Variable	Fish Richness (Sekong)			Fish Abundance (Sekong)			Fish Richness (Sre Pok)		
	Lag	Coefficient (Standard Error)	*p-value*	Lag	Coefficient (Standard Error)	*p-value*	Lag	Coefficient (Standard Error)	*p-value*
WL	--	--	--	11	-0.21 (0.093)	.026	9	-0.15 (0.079)	.065
T-P	--	--	--	9	0.31 (0.091)	.001	--	--	--
DO	--	--	--	--	--	--	0	-0.23 (0.079)	.005
NO_{2-3}-N	--	--	--	--	--	--	8	0.26 (0.080)	.002
NH_4-N	--	--	--	--	--	--	--	--	--
T	--	--	--	--	--	--	--	--	--
pH	11	-0.16 (0.095)	0.0959	--	--	--	--	--	--
Adjusted R^2	0.017			**0.115**			**0.159**		
p-value	0.096			**<0.001**			**<0.001**		

Variable	Fish Abundance (Sre Pok)			Fish Richness (Sesan)			Fish Abundance (Sesan)		
	Lag	Coefficient (Standard Error)	*p-value*	Lag	Coefficient (Standard Error)	*p-value*	Lag	Coefficient (Standard Error)	*p-value*
WL	10	-0.18 (0.100)	.077	--	--	--	--	--	--
T-P	4	0.30 (0.103)	.005	--	--	--	--	--	--
DO	--	--	--	--	--	--	--	--	--
NO_{2-3}-N	4	-0.25 (0.092)	.007	--	--	--	1	-0.19 (0.120)	.053
NH_4-N	--	--	--	4	-0.15 (0.047)	.002	0	-0.21 (0.109)	.121
T	--	--	--	6	-0.14 (0.052)	.009	--	--	--
pH	1	0.26 (0.091)	.005	--	--	--	--	--	--
Adjusted R^2		**0.219**			**0.174**			**0.073**	
p-value		**<0.001**			**<0.001**			**0.043**	

Table 5. *Multiple regression describing the influence of time-lagged hydrology to fish community before and after dam construction at Sekong and Sre Pok rivers. Variables are selected to maximize the p-value, and significant results are bolded.*

Variable	Fish Richness (Sekong) Before			After			Fish Abundance (Sekong) Before			After		
	Lag	Coefficient (Standard Error)	*p-value*	Lag	Coefficient (Standard Error)	*p-value*	Lag	Coefficient (Standard Error)	*p-value*	Lag	Coefficient (Standard Error)	*p-value*
WL	--	--	--	0	-0.12 (0.092)	.178	11	-0.13 (0.094)	0.185	--	--	--
T-P	--	--	--	--	--	--	--	--	--	--	--	--
DO	--	--	--	--	--	--	11	-0.15 (0.093)	0.107	--	--	--
NO_{2-3}-N	--	--	--	0	0.22 (0.099)	.028	--	--	--	2	-0.27 (0.098)	.006
NH_4-N	--	--	--	--	--	--	11	0.17 (0.089)	0.056	--	--	--
T	--	--	--	--	--	--	--	--	--	--	--	--
pH	10	-0.20 (0.095)	.040	3	-0.31 (0.093)	.001	--	--	--	--	--	--
Adjusted R^2		**0.030**			**0.114**			0.033			**0.061**	
p-value		**0.040**			**0.001**			0.086			**0.006**	

Variable	Fish Richness (Sre Pok) Before			After			Fish Abundance (Sre Pok) Before			After		
	Lag	Coefficient (Standard Error)	*p-value*	Lag	Coefficient (Standard Error)	*p-value*	Lag	Coefficient (Standard Error)	*p-value*	Lag	Coefficient (Standard Error)	*p-value*
WL	0	0.26 (0.0846)	.003	2	0.22 (0.091)	.020	--	--	---	--	--	--
	9	0.3647		--	--		4	0.36 (0.089)		8		.007
T-P		(0.0835)	<.001			--			<.001		0.26 (0.093)	
DO	--	--	--	0	-0.26 (0.091)	.005	--	--	--	--	--	--
NO_{2-3}-N	--	--	--	--	--	--	1	-0.33 (0.087)	<.001	--	--	--
NH_4-N	--	--	--	--	--	--	--	--	--	--	--	--
T	--	--	--	--	--	--	--	--	--	1	-0.20 (0.095)	.035
pH	--	--	--	--	--	--	--	--	--	--	--	--
Adjusted R^2		**0.163**			**0.120**			**0.177**			**0.068**	
p-value		**<0.001**			**<0.001**			**<0.001**			**0.009**	

4.4. Lagged time series regression before and after dam construction

All models except fish abundance before dam construction at Sekong show statistically significant results (Table 5). For instance, the lagged regression model of fish richness at Sekong over the whole study period is not statistically significant but both the pre-dam and post-dam models are, with significantly improved *p-values*. This may indicate that the dam modifies the dynamics between hydrology and fish, and consequently, separate models are better at explaining the changes in fish richness.

WL explains 4 out of the 8 models with three models showing immediate effects (0-2 months) and one model showing a long lag of 11 months. While WL positively correlates to fish abundance and richness at Sre Pok, it negatively correlates at Sekong. T-P explains 3 out of 8 models at lags of 4 to 9 months. Similar to the baseline and the combined lagged regression, T-P positively correlates to fish richness or abundance. DO negatively correlates with fish richness after dam construction at Sre Pok and with fish abundance before dam construction at Sekong, respectively. NO_{2-3}-N determines both fish richness and abundance at Sekong after dam construction and determines fish abundance at Sre Pok before dam construction, all at a short lag. pH correlates negatively to fish richness at Sekong at a lag of 10 months (pre-dam) and 3 months (post-dam). NH_4-N and T only each appear once in the selected models. Cocurrent with the increase and seasonal dynamic shifts in T at Sre Pok after dam construction (Fig. 4c), T negatively correlates to fish abundance post-dam.

5. Discussion

This study provides an assessment of the environmental losses of the hydropower dams and key hydrological parameters influencing fish abundance and richness. Regarding the first hypothesis, fish richness and abundance decrease with hydropower development, concurring with Ngor *et al.* (2017), except for fish abundance at Sekong. The large decrease in fish richness by 12% to 22% in the three rivers indicates that some endangered species or species highly reliant on certain habitats may disappear from this region (Ziv *et al.*, 2012). Fish richness at the most heavily disturbed river, Sesan, shows the lowest fish richness, concurring with Ou & Winemiller (2016). Corresponding to the alterations in fish, several hydrological parameters also shift before and after dam construction, which may lead to complex responses from the fish community (Alvarez-Mieles *et al.*, 2013; Fabricius *et al.*, 2004).

The seasonal dynamics of water level and water temperature decrease in magnitude after dam construction at most of the sites studied. Most of the migratory fish reside in the floodplain from July to November with high water levels. Water levels drawing down during this season decreases the amount of habitat available (Gaboury and Palatas, 1984), therefore reducing fish richness and abundance. The shrunken seasonal pattern in water temperature corresponds to findings by Bonnema *et al.* (2020) and Ou & Winemiller (2016). Fish are ectotherms, and thus the increase in temperature from July to November when adult fish and larvae move to tributary floodplains can lead to disruptions in homeostasis (Prosser & Nelson, 1981), changing metabolic rate (Killen *et al.*, 2010), feeding (Volkoff & Rønnestad, 2020), locomotion (Jahan, 2018), and reproductive ability (Soria *et al.*, 2008).

The second hypothesis is mostly supported, as two models in the baseline regression and five out of six models in the time-lagged regression display statistically significant results, indicating that a combination of time-lagged and immediate effects explains the change in fish richness and abundance. The conclusion corresponds to previous studies, demonstrating how heterogeneous habitats, combined with river characteristics and physicochemical properties define patterns of fish diversity (Kang *et al.*, 2009; Chea *et al.*, 2016). The parameters selected in the optimal model mostly correspond to variables showing significant changes before and after dam construction in Figure 3, showing that the decrease in fish populations occurs concurrently with changes in important hydrological factors. However, responses to the hydrological changes are complex and bidirectional, which

represents the multitude of ways fish communities adapt to vertical and horizontal habitat changes (Kramer, 1987).

The third hypothesis is supported, as the selected lagged variables are very different between the pre-dam and post-dam models, suggesting that dam construction also disrupts the patterns of interaction between hydrology and fish populations, likely caused by the shifts in hydrology. For example, at Sre Pok, following a 9.3% increase in dissolved oxygen, dissolved oxygen negatively correlates to fish richness post-dam construction, while not appearing as a significant factor in the pre-dam model. This can be either caused by a change in the fish assemblage structure due to shifting hydrology (e.g., Tan *et al.*, 2010; Baran *et al.*, 2011; Taylor *et al.*, 2014; Mims & Olden, 2012), with the new fish community having different environmental requirements (Stoffers *et al.*, 2020), or by changing levels that eliminate or add limiting factors (e.g., Ou & Winemiller, 2016). For instance, Ou & Winemiller (2016) found that primary production sources sustaining fish shift at Sesan due to hydrological changes; Mims & Olden (2012) found that reduced flow variability increased the proportion of equilibrium species and decreased opportunistic species. Furthermore, responses of fish to hydrological conditions in floodplain rivers are often density-dependent (Halls & Welcomme, 2004), and therefore differences in fish abundance may adjust such responses, altering the relationship between fish and hydrology. However, more research is needed to prove and explain this altered correlation between fish and hydrology.

All hydrological variables appear in at least one model in both combined and separate (pre-dam and post-dam) regressions, indicating that the fish community is closely connected to hydrological characteristics in the Mekong.

Water level occurs in most models for the three regression methods, agreeing with Poff (1997) that defines water level as a "master variable." While water level most closely corresponds to fish richness and abundance at 9 to 11 months lag in the combined model, separate models show a more immediate correlation between water level and fish with one exception. Due to the cyclic pattern and regular periodicity of water level in the 3S region (Piman *et al.*, 2016), the differing lag factors may represent two points of correspondence between the cycles of fish community and that of water level (Seeboonruang, 2014). Apart from Sre Pok, water level negatively correlates with fish richness and abundance, which matches other studies on sub-tropical floodplain rivers (Espínola *et al.*, 2016).

Unexpectedly, the dissolved oxygen at lags of 0 and 11 months negatively correlates with fish richness. Rising dissolved oxygen levels may cause supersaturation. Some fish species may develop antioxidant defenses and survive (Ross *et al.*, 2001) while others may perish, changing the fish assemblage structures in the three rivers, thus potentially modifying fish richness.

In all regression models, total phosphorous positively corresponds with fish richness and abundance and mainly controls fish abundance. Total phosphorous is a pivotal limiting factor determining the primary productivity in the waterbody (Jones & Lee, 1982), explaining the positive correlation. Furthermore, total phosphorous acts on the fish community by promoting the growth of algae (Smith & Kalff, 1981) over longer periods, which corresponds to the relatively longer lag of 4 to 9 months. However, while most species suffer in hypereutrophic lakes, *Cyprinidae*, which already dominates the 3S basin ecosystem (Montaña *et al.*, 2020), can increase in large quantities (Heminen *et al.*, 2000), leading to more dominance of the cyprinids (Tammi *et al.*, 2001). With increasing total phosphorous levels by almost 50% at Sekong and Sesan and the range of total phosphorous already exceeding the desired

concentration (Boyd, 2003), such changes can also reduce the balance between fish families and degrade richness.

Nitrite is a toxicant and a disruptor of multiple physiological functions in fish (Kroupova *et al.*, 2005). In Sekong and Sre Pok's lagged regression, fish richness positively corresponds yet fish abundance negatively corresponds to nitrite-nitrate. Depending on differential chloride ion uptake (Jensen, 2003), fish have differential tolerance levels to nitrite-nitrate in the water. Cyprinids are generally not tolerant to nitrite (Kroupova *et al.*, 2010; Kroupova *et al.*, 2006), but other fishes in the basin, such as *Chitala ornata* (Huong *et al.*, 2020), are not sensitive to these changes. Therefore, increasing nitrite decreases the dominating Cyprinidae, leading to lower fish quantities but allowing other species to thrive, thus increasing richness (cf. Duque *et al.*, 2020). Additionally, all except one lagged regression show a small lag (0 to 2 months), which reflects the immediate toxicity of nitrite-nitrate to fish (Jensen, 2004).

It is worthwhile to note that ammonium concentration is a strong influencing factor at Sesan that negatively correlates to both fish richness and abundance, while not a significant factor in the other two rivers. While ammonium (NH_4^+) itself is harmless, ammonia (NH_3) is highly toxic to fish (Korner *et al.*, 2001). Both substances can enter the freshwater system through fertilizer leaks (Aneja *et al.*, 2002). As the Sesan is the most populated and developed river in the basin (Phyrom *et al.*, 2013), extensive use of fertilizers in agriculture, one of the main forms of livelihood, may contribute to the increase in ammonium and ammonia concentration. Since water temperature and pH did not decrease at Sesan, this increase in ammonium would not be due to a change in equilibrium between ammonia and ammonium (Korner *et al.*, 2001; Johansson & Wedborg, 1980). Therefore, it is likely that the negative correlation between fish and ammonium may indicate the changes in ammonia that govern the fish communities.

Overall, the direction of change in the hydrological variables and the parameters used in the regression models differ significantly between rivers, which may be a result of individual riverine characteristics (Delong & Thoms, 2016) or differences in the human disturbance. For instance, the river shows a gradient of disturbance from Sesan to Sekong, where the population is mostly concentrated around Sesan, followed closely by Sre Pok, with a much lower density at Sekong (Phyrom *et al.*, 2013). Furthermore, the Sesan river has been heavily disturbed by the construction of three major dams since 2001 (Ngor, 2019), which is also shown to contribute to the heterogeneous responses (Ou & Winemiller, 2016).

5.1. Assessment of Methodology

The lagged multiple regression method accounts for the time differences in understanding how hydrological regimes affect fish. Model performance is significantly improved when using time-lagged variables, compared to the baseline scenario. However, the CCF regression may introduce problems of intra-multiplicity (Olden & Neff, 2001) and inferred coefficients (Orduz & Pickering, 2021), and many models exhibit similar correlation, confusing the determination of the optimal lag (Seeboonruang, 2014). Using time-lagged variables in multiple regression is a relatively uncommon approach, and more studies are needed to verify such a method and provide more supportive and robust conclusions.

Although using fisher-reported data with stationary gillnets may be a cost-effective approach for assessing the changes in the fish community over time, it also induces biases. The likelihood of fish being caught in the net decrease with increased

visibility and temperature (Hansson & Rudstam, 1995). Therefore, since water temperature increased at all sites and total suspended solids decreased at Sre Pok, the change in fish catchment may also be partly explained by the lower success rates.

Furthermore, the fishers are active in a variety of habitats in the 3S region, including rice fields, tributaries, and ponds, which host different fish communities (MRC, 2007). The influences of dams may drive fishers to fish in alternative habitats to maximize the number of fish caught, which may mitigate the impacts of hydropower to fish catchment data.

The research can also be expanded to provide more conclusive results. Annual precipitation may play an important role in driving aquatic biodiversity (Konar *et al.*, 2012), but it is not assessed due to limited data. Therefore, future studies should focus on uncovering relationships between hydrology and other environmental factors (Gunawardana *et al.*, 2021). Finally, due to a lack of fish assemblage analysis, conclusions are limited on whether the change in fish richness is due to decrease in dominance or other factors. Consequently, evaluations of assemblage structure before and after dam construction are needed.

Finally, the temporal changes in the hydrological regimes and fish are not just impacted by hydropower construction but also governed by climate change (Pokhrel *et al.*, 2017; Jacobson *et al.*, 2010), deforestation (Lohani *et al.*, 2020), and increased fishing pressure (Allan *et al.*, 2005). Therefore, studies should aim to quantify changes induced by hydropower developments by considering these variables in the analysis.

5.2. Implications to the fishery and the ecosystem

This research concludes that dam construction not only negatively affects fish communities but also changes the dynamics between fish and hydrology. Yet, the continued dam construction in the 3S basin also impacts the whole aquatic ecosystem and the livelihoods of the villagers.

As an indicator of ecosystem health, fish are the keystone of an aquatic ecosystem, and their population changes epitomize the changes in river health (Fausch *et al.*, 1990; Harris, 1995). Decreased abundance and richness in fish, therefore, may signalize a greater degradation in the aquatic ecosystem in the region, leading to cascades of effects and irrecoverable loss in one of the richest regions of the Lower Mekong basin (Moyle & Leidy, 1992).

Furthermore, the decrease in fish abundance and long-term degradation of ecosystem functions pose threats to the poverty-stricken residents. Residents in the Lower Mekong Basin have long depended on fish as their main intake of meat and protein, more than any major basin in the world (Hortle, 2007), and 66% of them are involved in fisheries (MRC, 2010). At the same time, the 3S region remains the least developed in terms of socio-economic status across Cambodia, and only 10% of the population has access to electricity in the basin (Phyrom *et al.*, 2013). While hydropower stations provide electricity to address local needs (Grumbine & Xu, 2011), they also deteriorate fisheries (Orr *et al.*, 2012) and dislocate villagers (Manorom, 2018), leading to more social issues.

5.3. Conclusion

Resolving the dilemma between human development and environmental protection requires both scientific contribution and governmental devotion (Kummu *et al.*, 2006), and the problem is complicated by the fact that the residents can be adversely

affected in both scenarios. While Cambodia has halted dam construction on the main river (Kijewski, 2020), legal frameworks and actions are still lacking on the tributaries. Therefore, this study calls for a re-evaluation of the dam systems in the tributaries of Mekong and advocates for more research on the interactions between the river, ecosystem, and people to make a comprehensive suggestion on the tradeoffs between the fish and the dams.

References

Adler, D. & Kelly, T. S. (2021). vioplot: violin plot. R package version 0.3.7 https://github.com/TomKellyGenetics/vioplot

Allan, J. D., Abell, R., Hogan, Z., Revenga, C., Taylor, B. W., Welcomme, R. L., & Winemiller, K. (2005). Overfishing of Inland Waters. *BioScience*, *55*(12), 1041. https://doi.org/10.1641/0006-3568(2005)055[1041:OOIW]2.0.CO;2

Altermatt, F., & Fronhofer, E. A. (2018). Dispersal in dendritic networks: Ecological consequences on the spatial distribution of population densities. *Freshwater Biology*, *63*(1), 22–32. https://doi.org/10.1111/fwb.12951

Alvarez-Mieles, G., Irvine, K., Griensven, A. V., Arias-Hidalgo, M., Torres, A., & Mynett, A. E. (2013). Relationships between aquatic biotic communities and water quality in a tropical river–wetland system (Ecuador). *Environmental Science & Policy*, *34*, 115–127. https://doi.org/10.1016/j.envsci.2013.01.011

Aneja, V. P., Nelson, D. R., Roelle, P. A., & Walker, J. T. (2003). Agricultural ammonia emissions and ammonium concentrations associated with aerosols and precipitation in the Southeast United States. *Journal of Geophysical Research*, *108*(4). https://doi.org/10.1029/2002jd002271

Baird, I. G. (2014, August 9). *Cambodia's LS2 Dam is a disaster in the making* . East Asia Forum. Retrieved July 5, 2022, from https://www.eastasiaforum.org/2014/08/09/cambodias-ls2-dam-is-a-disaster-in-the-making/

Baran, E., Guerin, E., & Nasielski, J. (2015). *Fish, sediment, and dams in the Mekong*. WorldFish, and CGIAR Research Program on Water, Land and Ecosystems.

Basavaraja, D., Narayana, J., Kiran, B. R., & Puttaiah, E. T. (2014). *Fish diversity and abundance in relation to water quality of Anjanapura reservoir, Karnataka, India*. 11.

Bennett, R.J., Spatial Time Series. London: Pion Limited, 1979.

Boyd, C. E. (2003). Guidelines for aquaculture effluent management at the farm-level. *Aquaculture*, *226*(1–4), 101–112. https://doi.org/10.1016/S0044-8486(03)00471-X

Bunthang, T., & Phen, C. (2021). *Fish Spawning Habitats in the Mekong and 3S Rivers in Cambodia*. 5th APBON Web Seminar, Cambodia. http://www.esabii.biodic.go.jp/ap-bon/meetings/documents/webseminar2020/0121_1.pdf

Chantha, O., & Ty, S. (2020). Assessing changes in flow and water quality emerging from hydropower development and operation in the Sesan River Basin of the Lower Mekong Region. *Sustainable Water Resources Management*, *6*(2), 27. https://doi.org/10.1007/s40899-020-00386-8

Chea, R., Grenouillet, G., & Lek, S. (2016). Evidence of Water Quality Degradation in Lower Mekong Basin Revealed by Self-Organizing Map. *PLOS ONE*, *11*(1), e0145527. https://doi.org/10.1371/journal.pone.0145527

Chea, R., Lek, S., Ngor, P., & Grenouillet, G. (2017). Large-scale patterns of fish diversity and assemblage structure in the longest tropical river in Asia. *Ecology of Freshwater Fish*, *26*(4), 575–585. https://doi.org/10.1111/eff.12301

Constable, D. (2015) The Sesan and Srepok River Basins. Bangkok, Thailand, IUCN. 56pp.

Delong, M. D., & Thoms, M. C. (2016). Changes in the trophic status of fish feeding guilds in response to flow modification: Trophic Responses to Flow Change. *Journal of Geophysical Research: Biogeosciences*, *121*(3), 949–964. https://doi.org/10.1002/2015JG003249

Duque, G., Gamboa-García, D. E., Molina, A., & Cogua, P. (2020). Effect of water quality variation on fish assemblages in an anthropogenically impacted tropical estuary, Colombian Pacific. *5*

Eslami, S., Hoekstra, P., Trung, N., & Kantoush, S. A. (n.d.). *OPEN Tidal amplification and salt*. 11.

Espínola, L. A., Rabuffetti, A. P., Abrial, E., Amsler, M. L., Blettler, M. C. A., Paira, A. R., Simões, N. R., & Santos, L. N. (2017). Response of fish assemblage structure to changing flood and flow pulses in a large subtropical river. *Marine and Freshwater Research*, *68*(2), 319. https://doi.org/10.1071/MF15141

Eyler, B. (2021, April 30). 2020 status of Lower Mekong Mainstream and tributary dams. Stimson Center. Retrieved June 7, 2022, from https://www.stimson.org/2020/2020-status-of-lower-mekong-mainstream-and-tributary-dams/

Fabricius, K., De'ath, G., McCook, L., Turak, E., & Williams, D. McB. (2005). Changes in algal, coral and fish assemblages along water quality gradients on the inshore Great Barrier Reef. *Marine Pollution Bulletin*, *51*(1–4), 384–398. https://doi.org/10.1016/j.marpolbul.2004.10.041

Fan, H., He, D., & Wang, H. (2015). Environmental consequences of damming the mainstream Lancang-Mekong River: A review. *Earth-Science Reviews*, *146*, 77–91. https://doi.org/10.1016/j.earscirev.2015.03.007

Fantin-Cruz, I., Pedrollo, O., Girard, P., Zeilhofer, P., & Hamilton, S. K. (2016). Changes in river water quality caused by a diversion hydropower dam bordering the Pantanal floodplain. *Hydrobiologia*, *768*(1), 223–238. https://doi.org/10.1007/s10750-015-2550-4

Fausch, K. D., Lyons, J., Karr, J. R., & Angermeier, P. L. (1990). Fish Communities as Indicators of Environmental Degradation. *American Fisheries Society*, *8*, 123–144.

Fornaroli, R., Muñoz-Mas, R., & Martínez-Capel, F. (2020). Fish community responses to antecedent hydrological conditions based on long-term data in Mediterranean river basins (Iberian Peninsula). *Science of The Total Environment*, *728*, 138052. https://doi.org/10.1016/j.scitotenv.2020.138052

Gaboury, M. N., & Patalas, J. W. (1984). Influence of Water Level Drawdown on the Fish Populations of Cross Lake, Manitoba. *Canadian Journal of Fisheries and Aquatic Sciences*, *41*(1), 118–125. https://doi.org/10.1139/f84-011

Ganassin, M. J. M., Muñoz-Mas, R., de Oliveira, F. J. M., Muniz, C. M., dos Santos, N. C. L., García-Berthou, E., & Gomes, L. C. (2021). Effects of reservoir cascades on diversity, distribution, and abundance of fish assemblages in three Neotropical basins. *Science of The Total Environment*, *778*, 146246. https://doi.org/10.1016/j.scitotenv.2021.146246

Grill, G., Ouellet Dallaire, C., Fluet Chouinard, E., Sindorf, N., & Lehner, B. (2014). Development of new indicators to evaluate river fragmentation and flow regulation at large scales: A case study for the Mekong River Basin. *Ecological Indicators*, *45*, 148–159. https://doi.org/10.1016/j.ecolind.2014.03.026

Grumbine, R. E., & Xu, J. (2011). Mekong Hydropower Development. *Science*, *332*(6026), 178–179. https://doi.org/10.1126/science.1200990

Gunawardana, S. K., Shrestha, S., Mohanasundaram, S., Salin, K. R., & Piman, T. (2021). Multiple drivers of hydrological alteration in the transboundary Srepok River Basin of the Lower Mekong Region. *Journal of Environmental Management*, *278*, 111524. https://doi.org/10.1016/j.jenvman.2020.111524

Halls, A. S., & Welcomme, R. L. (2004). Dynamics of river fish populations in response to hydrological conditions: A simulation study. *River Research and Applications*, *20*(8), 985–1000. https://doi.org/10.1002/rra.804

Hansson, S., & Rudstam, L. G. (1995). Gillnet catches as an estimate of fish abundance: A comparison between vertical gillnet catches and hydroacoustic abundances of Baltic Sea herring (*Clupea harengus*) and sprat (*Sptattus sptattus*). *Canadian Journal of Fisheries and Aquatic Sciences*, *52*(1), 75–83. https://doi.org/10.1139/f95-007

Harris, J. H. (1995). The use of fish in ecological assessments. *Austral Ecology*, *20*(1), 65–80. https://doi.org/10.1111/j.1442-9993.1995.tb00523.x

Hecht, J. S., Lacombe, G., Arias, M. E., Dang, T. D., & Piman, T. (2019). Hydropower dams of the Mekong River basin: A review of their hydrological impacts. *Journal of Hydrology*, *568*, 285–300. https://doi.org/10.1016/j.jhydrol.2018.10.045

Hogan, Z., Baird, I. G., Radtke, R., & Vander Zanden, M. J. (2007). Long distance migration and marine habitation in the tropical Asian catfish, Pangasius krempfi. *Journal of Fish Biology*, *71*(3), 818–832. https://doi.org/10.1111/j.1095-8649.2007.01549.x

Hortle, K.G. (2007). Consumption and the yield of fish and other aquatic animals from the Lower Mekong Basin, MRC Technical Paper No. 16. Mekong River Commission, Vientiane.

Huang, F., Xia, Z., Zhang, N., & Lu, Z. (2011). Does hydrologic regime affect fish diversity? -A case study of the Yangtze Basin (China). *Environmental Biology of Fishes*, *92*(4), 569–584. https://doi.org/10.1007/s10641-011-9880-5

Huong, D. T. T., Gam, L. T. H., Lek, S., Ut, V. N., & Phuong, N. T. (2020). Effects of nitrite at different temperatures on physiological parameters and growth in clown knifefish (Chitala ornata, Gray 1831). *Aquaculture*, *521*, 735060. https://doi.org/10.1016/j.aquaculture.2020.735060

Jacobson, P. C., Stefan, H. G., & Pereira, D. L. (2010). Coldwater fish oxythermal habitat in Minnesota lakes: Influence of total phosphorus, July air temperature, and relative depth. *Canadian Journal of Fisheries and Aquatic Sciences*, *67*(12), 2002–2013. https://doi.org/10.1139/F10-115

Jahan, I. (2018). *Impact of Temperature Increase on Freshwater Fish Species: Energetics and Muscle Mechanics of Two Centrarchids* [Easter Illionis University]. https://thekeep.eiu.edu/theses/4470

Jensen, F. B. (2003). Nitrite disrupts multiple physiological functions in aquatic animals. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, *135*(1), 9–24. https://doi.org/10.1016/S1095-6433(02)00323-9

Johansson, O., & Wedborg, M. (1980). The ammonia-ammonium equilibrium in seawater at temperatures between 5 and 25°C. *Journal of Solution Chemistry*, *9*(1), 37–44. https://doi.org/10.1007/bf00650135

Jones, R. A., & Lee, G. F. (1982). Recent advances in assessing impact of phosphorus loads on eutrophication-related water quality. *Water Research*, *16*(5), 503–515. https://doi.org/10.1016/0043-1354(82)90069-0

Kang, B., He, D., Perrett, L., Wang, H., Hu, W., Deng, W., & Wu, Y. (2009). Fish and fisheries in the Upper Mekong: Current assessment of the fish community, threats and conservation. *Reviews in Fish Biology and Fisheries*, *19*(4), 465–480. https://doi.org/10.1007/s11160-009-9114-5

Kijewski, L. (2020, April 1). *Cambodia halts hydropower construction on Mekong River until 2030*. VOA. Retrieved July 5, 2022, from https://www.voanews.com/a/east-asia-pacific_cambodia-halts-hydropower-construction-mekong-river-until-203/6186756.html

Killen, S. S., Atkinson, D., & Glazier, D. S. (2010). The intraspecific scaling of metabolic rate with body mass in fishes depends on lifestyle and temperature. *Ecology Letters*, *13*(2), 184–193. https://doi.org/10.1111/j.1461-0248.2009.01415.x

Konar, M., Jason Todd, M., Muneepeerakul, R., Rinaldo, A., & Rodriguez-Iturbe, I. (2013). Hydrology as a driver of biodiversity: Controls on carrying capacity, niche formation, and dispersal. *Advances in Water Resources*, *51*, 317–325. https://doi.org/10.1016/j.advwatres.2012.02.009

Kondolf, G. M., Rubin, Z. K., & Minear, J. T. (2014). Dams on the Mekong: Cumulative sediment starvation. *Water Resources Research*, *50*(6), 5158–5169. https://doi.org/10.1002/2013WR014651

Körner, S., Das, S. K., Veenstra, S., & Vermaat, J. E. (2001). The effect of ph variation at the ammonium/ammonia equilibrium in wastewater and its toxicity to Lemna Gibba. *Aquatic Botany*, *71*(1), 71–78. https://doi.org/10.1016/s0304-3770(01)00158-9

Kramer, D. L. (1987). Dissolved oxygen and fish behavior. *Environmental Biology of Fishes*, *18*(2), 81–92. https://doi.org/10.1007/BF00002597

Kroupova, H., Machova, J., & Svobodova, Z. (2005). Nitrite influence on fish: A review. *Veterinární Medicína*, *50*(No. 11), 461–471. https://doi.org/10.17221/5650-VETMED

Kroupová, H., Máchová, J., Piačková, V., Flajšhans, M., Svobodová, Z., & Poleszczuk, G. (2006). Nitrite Intoxication of Common Carp (Cyprinus carpio L.) at Different Water Temperatures. *Acta Veterinaria Brno*, *75*(4), 561–569. https://doi.org/10.2754/avb200675040561

Kroupova, H., Prokes, M., Macova, S., Penaz, M., Barus, V., Novotny, L., & Machova, J. (2010). Effect of nitrite on early-life stages of common carp (Cyprinus carpio L.). *Environmental toxicology and chemistry*, *29*(3), 535–540. https://doi.org/10.1002/etc.84

Kummu, M., & Varis, O. (2007). Sediment-related impacts due to upstream reservoir trapping, the Lower Mekong River. *Geomorphology*, *85*(3–4), 275–293. https://doi.org/10.1016/j.geomorph.2006.03.024

Kummu, M., Sarkkula, J., Koponen, J., & Nikula, J. (2006). Ecosystem Management of the Tonle Sap Lake: An Integrated Modelling Approach. *International Journal of Water Resources Development*, *22*(3), 497–519. https://doi.org/10.1080/07900620500482915

Li, J., Dong, S., Peng, M., Yang, Z., Liu, S., Li, X., & Zhao, C. (2013). Effects of damming on the biological integrity of fish assemblages in the Middle Lancang-Mekong River Basin. *Ecological Indicators*, *34*, 94–102. https://doi.org/10.1016/j.ecolind.2013.04.016

Lohani, S., Dilts, T., Weisberg, P., Null, S., & Hogan, Z. (2020). Rapidly Accelerating Deforestation in Cambodia's Mekong River Basin: A Comparative Analysis of

Spatial Patterns and Drivers. *Water*, *12*(8), 2191. https://doi.org/10.3390/w12082191

Lu, X. X., & Chua, S. D. X. (2021). River Discharge and Water Level Changes in the Mekong River: Droughts in an Era of MEGA-DAMS. *Hydrological Processes*, *35*(7). https://doi.org/10.1002/hyp.14265

Manh, N. V., Dung, N. V., Hung, N. N., Kummu, M., Merz, B., & Apel, H. (2015). Future sediment dynamics in the Mekong Delta floodplains: Impacts of hydropower development, climate change and sea level rise. *Global and Planetary Change*, *127*, 22–33. https://doi.org/10.1016/j.gloplacha.2015.01.001

Manorom, K. (n.d.). *Hydropower Resettlement in the Mekong Region*. 16.

Mekong River Commision. (2014). *Summary of IKMP & WWF Sediment Investigations* (Summary Report of Decision Support for Generating Sustainable Hydropower in the Mekong Basin). Mekong River Commission Secretariat.

Mekong River Commision. (2017). *2015 Lower Mekong Regional Water Quality Monitoring Report*. Mekong River Commission Secretariat.

Mekong River Commission (2010). State of the Basin Report 2010. Mekong River Commission, Vientiane, Lao PDR.

Mekong River Commission. (2021). *Status and Trends of Fish Abundance and Diversity in the Lower Mekong Basin during 2007–2018*. Mekong River Commission Secretariat. https://doi.org/10.52107/mrc.qx5yo0

Mims, M. C., & Olden, J. D. (2013). Fish assemblages respond to altered flow regimes via ecological filtering of life history strategies: *Fish assemblages respond to altered flow regimes*. *Freshwater Biology*, *58*(1), 50–62. https://doi.org/10.1111/fwb.12037

Mittermeier, R. A., Turner, W. R., Larsen, F. W., Brooks, T. M., & Gascon, C. (2011). Global Biodiversity Conservation: The critical role of hotspots. *Biodiversity Hotspots*, 3–22. https://doi.org/10.1007/978-3-642-20992-5_1

Montaña, C. G., Ou, C., Keppeler, F. W., & Winemiller, K. O. (2020). Functional and trophic diversity of fishes in the Mekong-3S river system: Comparison of morphological and isotopic patterns. *Environmental Biology of Fishes*, *103*(2), 185–200. https://doi.org/10.1007/s10641-020-00947-y

Moyle, P. B., & Leidy, R. A. (1992). Loss of Biodiversity in Aquatic Ecosystems: Evidence from Fish Faunas. In P. L. Fiedler & S. K. Jain (Eds.), *Conservation Biology* (pp. 127–169). Springer US. https://doi.org/10.1007/978-1-4684-6426-9_6

Moyle, P. B., & Leidy, R. A. (1992). *Loss of Biodiversity in Aquatic Ecosystems: Evidence from Fish Faunas. Conservation Biology, 127–169*. doi:10.1007/978-1-4684-6426-9_6

MRC (2007). Monitoring fish abundance and diversity in the Lower Mekong Basin: methodological guidelines. Mekong River Commission, Phnom Penh, Cambodia.

Ngor, P. B. (2018). *Fish assemblages dynamic in the tropical flood-pulse system of the Lower Mekong River Basin* [Unpublished]. http://rgdoi.net/10.13140/RG.2.2.16505.11368

Ngor, P. B., Legendre, P., Oberdorff, T., & Lek, S. (2018). Flow alterations by dams shaped fish assemblage dynamics in the complex Mekong-3S river system. *Ecological Indicators*, *88*, 103–114. https://doi.org/10.1016/j.ecolind.2018.01.023

Ngor, P. B., Oberdorff, T., Phen, C., Baehr, C., Grenouillet, G., & Lek, S. (2018). Fish assemblage responses to flow seasonality and predictability in a tropical flood pulse system. *Ecosphere*, *9*(11), e02366. https://doi.org/10.1002/ecs2.2366

Nilsson, C., Ekblad, A., Gardfjell, M., & Carlberg, B. (1991). Long-Term Effects of River Regulation on River Margin Vegetation. *The Journal of Applied Ecology*, *28*(3), 963. https://doi.org/10.2307/2404220

Null, S. E., Farshid, A., Goodrum, G., Gray, C. A., Lohani, S., Morrisett, C. N., Prudencio, L., & Sor, R. (2020). A Meta-Analysis of Environmental Tradeoffs of Hydropower Dams in the Sekong, Sesan, and Srepok (3S) Rivers of the Lower Mekong Basin. *Water*, *13*(1), 63. https://doi.org/10.3390/w13010063

Nuon, V., Lek, S., Ngor, P. B., So, N., & Grenouillet, G. (2020). Fish Community Responses to Human-Induced Stresses in the Lower Mekong Basin. *Water*, *12*(12), 3522. https://doi.org/10.3390/w12123522

Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Solymos P, Stevens M, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S, Evangelista H, FitzJohn R, Friendly M, Furneaux B, Hannigan G, Hill M, Lahti L, McGlinn D, Ouellette M, Ribeiro Cunha E, Smith T, Stier A, Ter Braak C, Weedon J (2022). vegan: Community Ecology Package_. R package version 2.6-2, https://CRAN.R-project.org/package=vegan.

Olden, J. D., & Neff, B. D. (2001). Cross-correlation bias in lag analysis of aquatic time series. *Marine Biology*, *138*(5), 1063–1070. https://doi.org/10.1007/s002270000517

Orduz, J. C., & Pickering, A. (2021). *Modelling stochastic time delay for regression analysis* (arXiv:2111.06403). arXiv. http://arxiv.org/abs/2111.06403

Orr, S., Pittock, J., Chapagain, A., & Dumaresq, D. (2012). Dams on the Mekong River: Lost fish protein and the implications for land and water resources. *Global Environmental Change*, *22*(4), 925–932. https://doi.org/10.1016/j.gloenvcha.2012.06.002

Ou, C., & Winemiller, K. O. (2016). Seasonal hydrology shifts production sources supporting fishes in rivers of the Lower Mekong Basin. *Canadian Journal of Fisheries and Aquatic Sciences*, *73*(9), 1342–1362. https://doi.org/10.1139/cjfas-2015-0214

Phyrom, S., Keartha, C., Seng, S., Sajor, E., & Ongsakul, R. (n.d.). *Development of Water Resource Infrastructures and Livelihood Benefits: A Case of Lower Sesan 2 Project, Cambodia*. 28.

Piman, T., Cochrane, T. A., & Arias, M. E. (2016). Effect of Proposed Large Dams on Water Flows and Hydropower Production in the Sekong, Sesan and Srepok Rivers of the Mekong Basin: Impact of Large Dams in Tributaries of the Mekong. *River Research and Applications*, *32*(10), 2095–2108. https://doi.org/10.1002/rra.3045

Piman, T., Cochrane, T. A., Arias, M. E., Green, A., & Dat, N. D. (2013). Assessment of Flow Changes from Hydropower Development and Operations in Sekong, Sesan, and Srepok Rivers of the Mekong Basin. *Journal of Water Resources Planning and Management*, *139*(6), 723–732. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000286

Poff, N. L., & Ward, J. V. (1989). Implications of Streamflow Variability and Predictability for Lotic Community Structure: A Regional Analysis of Streamflow Patterns. *Canadian Journal of Fisheries and Aquatic Sciences*, *46*(10), 1805–1818. https://doi.org/10.1139/f89-228

Poff, N. L., Allan, J. D., Bain, M. B., Karr, J. R., Prestegaard, K. L., Richter, B. D., Sparks, R. E., & Stromberg, J. C. (1997). The Natural Flow Regime. *BioScience*, *47*(11), 769–784. https://doi.org/10.2307/1313099

Pokhrel, Y., Burbano, M., Roush, J., Kang, H., Sridhar, V., & Hyndman, D. (2018). A Review of the Integrated Effects of Changing Climate, Land Use, and Dams on Mekong River Hydrology. *Water*, *10*(3), 266. https://doi.org/10.3390/w10030266

Poulsen, A. F., Hortle, K. G., Valbo-Jorgensen, J., Chan, S., Chhuon, C. K., Viravong, S., Bouakhamvongsa, K., Suntornratana, U., Yoorong, N., Nguyen, T. T., Tran, B. Q., Hortle, E. K. G., Booth, S. J., & Visser, T. A. M. (2004). *Distribution and Ecology of Some Important Riverine Fish Species of the Mekong River Basin*. 116.

Prosser, C. L., & Nelson, D. O. (1981). The Role of Nervous Systems in Temperature Adaptation of Poikilotherms. *Annual Review of Physiology*, *43*(1), 281–300. https://doi.org/10.1146/annurev.ph.43.030181.001433

Pyron, M., Lauer, T. E., & Gammon, J. R. (2006). Stability of the Wabash River fish assemblages from 1974 to 1998. *Freshwater Biology*, *51*(10), 1789–1797. https://doi.org/10.1111/j.1365-2427.2006.01609.x

QGIS Development Team (2022). QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org".

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Räsänen, T. A., Joffre, O. M., Someth, P., Thanh, C. T., Keskinen, M., & Kummu, M. (2015). Model-Based Assessment of Water, Food, and Energy Trade-Offs in a Cascade of Multipurpose Reservoirs: Case Study of the Sesan Tributary of the Mekong River. *Journal of Water Resources Planning and Management*, *141*(1), 05014007. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000459

Ross, S. W., Dalton, D. A., Kramer, S., & Christensen, B. L. (2001). *Physiological ž antioxidant/ responses of estuarine fishes to variability in dissolved oxygen*. 15.

Schindler, D. W. (1974). Eutrophication and Recovery in Experimental Lakes: Implications for Lake Management. *Science*, *184*(4139), 897–899. https://doi.org/10.1126/science.184.4139.897

Seeboonruang, U. (2015). An application of time-lag regression technique for assessment of groundwater fluctuations in a regulated river basin: A case study in Northeastern Thailand. *Environmental Earth Sciences*, *73*(10), 6511–6523. https://doi.org/10.1007/s12665-014-3872-7

Shao, X., Fang, Y., Jawitz, J. W., Yan, J., & Cui, B. (2019). River network connectivity and fish diversity. *Science of The Total Environment*, *689*, 21–30. https://doi.org/10.1016/j.scitotenv.2019.06.340

Signs, M., Venturini, S., & Yoganand, K. (2021). *New species discoveries in the Greater Mekong 2020*. World Wide Fund for Nature. https://files.worldwildlife.org/wwfcmsprod/files/Publication/file/5s8a16akht_WWF_New_species_discoveries_2020_PAGES_final_compressed.pdf?_ga=2.51206753.91517555.1654599688-1784951956.1654506046

Smith, R. E. H., & Kalff, J. (1981). The Effect of Phosphorus Limitation on Algal Growth Rates: Evidence from Alkaline Phosphatase. *Canadian Journal of Fisheries and Aquatic Sciences*, *38*(11), 1421–1427. https://doi.org/10.1139/f81-188

Sor, R., Ngor, P. B., Soum, S., Chandra, S., Hogan, Z. S., & Null, S. E. (2021). Water Quality Degradation in the Lower Mekong Basin. *Water*, *13*(11), 1555. https://doi.org/10.3390/w13111555

Sor, R., Ngor, P., Boets, P., Goethals, P., Lek, S., Hogan, Z., & Park, Y.-S. (2020). Patterns of Mekong Mollusc Biodiversity: Identification of Emerging Threats and Importance to Management and Livelihoods in a Region of Globally Significant Biodiversity and Endemism. *Water*, *12*(9), 2619. https://doi.org/10.3390/w12092619

Soria, F. N., Strüssmann, C. A., & Miranda, L. A. (2008). High Water Temperatures Impair the Reproductive Ability of the Pejerrey Fish *Odontesthes bonariensis*: Effects on the Hypophyseal-Gonadal Axis. *Physiological and Biochemical Zoology*, *81*(6), 898–905. https://doi.org/10.1086/588178

Soukhaphon, A., Baird, I. G., & Hogan, Z. S. (2021). The Impacts of Hydropower Dams in the Mekong River Basin: A Review. *Water*, *13*(3), 265. https://doi.org/10.3390/w13030265

Sparks, R. E. (1995). Need for Ecosystem Management of Large Rivers and Their Floodplains. *BioScience*, *45*(3), 168–182. https://doi.org/10.2307/1312556

Stoffer, D. & Poison, N. (2022). *astsa: Applied Statistical Time Series Analysis*. R package version 1.15, https://CRAN.R-project.org/package=astsa.

Stoffers, T. (2022). Freshwater fish biodiversity restoration in floodplain rivers requires connectivity and habitat heterogeneity at multiple spatial scales. *Science of the Total Environment*, 12.

Tammi, J., Lappalainen, A., Mannio, J., Rask, M., & Vuorenmaa, J. (1999). Effects of eutrophication on fish and fisheries in Finnish lakes: A survey based on random sampling. *Fisheries Management and Ecology*, *6*(3), 173–186. https://doi.org/10.1046/j.1365-2400.1999.00152.x

Tan, X., Li, X., Lek, S., Li, Y., Wang, C., Li, J., & Luo, J. (2010). Annual dynamics of the abundance of fish larvae and its relationship with hydrological variation in the Pearl River. *Environmental Biology of Fishes*, *88*(3), 217–225. https://doi.org/10.1007/s10641-010-9632-y

Taylor, J. M., Seilheimer, T. S., & Fisher, W. L. (2014). Downstream fish assemblage response to river impoundment varies with degree of hydrologic alteration. *Hydrobiologia*, *728*(1), 23–39. https://doi.org/10.1007/s10750-013-1797-x

Tranmer, M., Murphy, J., Elliot, M., & Pampaka, M. (n.d.). *Multiple Linear Regression (2nd Edition)*. 59.

Tranmer, M., Murphy, J., Elliot, M., and Pampaka, M. (2020) Multiple Linear Regression (2nd Edition); Cathie Marsh Institute Working Paper 2020-01. https://hummedia.manchester.ac.uk/institutes/cmist/archive-publications/working-papers/2020/2020-1- multiple-linear-regression.pdf

Trung, L. D., Duc, N. A., Nguyen, L. T., Thai, T. H., Khan, A., Rautenstrauch, K., & Schmidt, C. (2020). Assessing cumulative impacts of the proposed Lower Mekong Basin hydropower cascade on the Mekong River floodplains and Delta – Overview of integrated modeling methods and results. *Journal of Hydrology*, *581*, 122511. https://doi.org/10.1016/j.jhydrol.2018.01.029

Valbo-Jørgensen, J., Coates, D., & Hortle, K. (2009). Fish Diversity in the Mekong River Basin. In *The Mekong* (pp. 161–196). Elsevier. https://doi.org/10.1016/B978-0-12-374026-7.00008-5

Volkoff, H., & Rønnestad, I. (2020). Effects of temperature on feeding and digestive processes in fish. *Temperature*, *7*(4), 307–320. https://doi.org/10.1080/23328940.2020.1765950

Walling, D. E. (2009). The Sediment Load of the Mekong River. In *The Mekong* (pp. 113–142). Elsevier. https://doi.org/10.1016/B978-0-12-374026-7.00006-1

Webber, M., Edwards-Myers, E., Campbell, C., & Webber, D. (2005). Phytoplankton and zooplankton as indicators of water quality in Discovery Bay, Jamaica. *Hydrobiologia*, *545*(1), 177–193. https://doi.org/10.1007/s10750-005-2676-x

Wei, G., Yang, Z., Cui, B., Li, B., Chen, H., Bai, J., & Dong, S. (2009). Impact of Dam Construction on Water Quality and Water Self-Purification Capacity of the Lancang River, China. *Water Resources Management*, *23*(9), 1763–1780. https://doi.org/10.1007/s11269-008-9351-8

Wickham, H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Wild, T. B., & Loucks, D. P. (2014). Managing flow, sediment, and hydropower regimes in the Sre Pok, Se San, and Se Kong Rivers of the Mekong basin. *Water Resources Research*, *50*(6), 5141–5157. https://doi.org/10.1002/2014WR015457

Winemiller, K. O., McIntyre, P. B., Castello, L., Fluet-Chouinard, E., Giarrizzo, T., Nam, S., Baird, I. G., Darwall, W., Lujan, N. K., Harrison, I., Stiassny, M. L. J., Silvano, R. A. M., Fitzgerald, D. B., Pelicice, F. M., Agostinho, A. A., Gomes, L. C., Albert, J. S., Baran, E., Jr, M. P., … Sáenz, L. (n.d.). *Balancing hydropower and biodiversity in the Amazon, Congo, and Mekong*. 2.

Wolman, M. G., & Miller, J. P. (1960). Magnitude and Frequency of Forces in Geomorphic Processes. *The Journal of Geology*, *68*(1), 54–74. https://doi.org/10.1086/626637

Xue, Z., Liu, J. P., & Ge, Q. (2010). *Changes in hydrology and sediment delivery of the Mekong River in the last 50 years: Connection to damming, monsoon, and ENSO*. *36*, 13.

Ziv, G., Baran, E., Nam, S., Rodríguez-Iturbe, I., & Levin, S. A. (2012). Trading-off fish biodiversity, food security, and hydropower in the Mekong River Basin. *Proceedings of the National Academy of Sciences*, *109*(15), 5609–5614. https://doi.org/10.1073/pnas.1201423109

Supply-Sided and Demand-Sided Solutions to Fast-Fashion's Social Impacts

Jeongho Ha

Author Background: *Jeongho Ha grew up in South Korea and currently attends Korea International School in Gyeonggi-do, South Korea. His Pioneer research concentration was in the field of environmental studies/economics and titled "Sustainable Development."*

Abstract

Fast-fashion is a term that describes a company implementing a volume-centric business model that profits by selling large quantities of cheap, readymade garments that are disposed of after a season. Though in itself problematic and conducive to negative social impacts such as microplastic pollution, the recent rise of fast-fashion-sponsored online influencers that embrace and even promote the practice of 'hauling', buying excessive amounts of cheap garments only to throw them away at the end of the season, have exacerbated the consequences of this model. With the fast-fashion industry's current position as the leading cause behind child labor in low income nations, the largest source of microplastic pollution in oceans, and the reason behind the downfall of domestic garment and textile industries in African nations, the need to find a solution to curb the social impacts of fast-fashion has become greater than ever. This paper attempted to propose feasible supply-sided and demand-sided solutions to this issue by first analyzing each of fast fashion's social impacts in detail and then formulating solutions by referring to methods that have seen success in a similar context or issue. Specifically, two solutions were devised for each solution type. It was determined that composition-based environmental taxes and a cradle-to-grave assessment of garments were effective supply-sided solutions, while an intuitive grading system for garments and the utilization of public figures to discourage fast-fashion consumption were projected to be effective demand-sided solutions.

1. Introduction

The word fast-fashion refers to a clothing company or a brand whose profit model is volume-focused and produces cheap and disposable clothing catering to the preferences of the youth. Initially coined by the *New York Times*, the term that was first used to describe the fashion label Zara's success in the 1990s in implementing its revolutionary 15 day design-to-production cycle (Rauturier, 2022) now sees ubiquitous use in labeling popular clothing brands. These fast-fashion brands now account for over 66% of the clothing market, and the world has seen a doubling of the total volume of garments produced in the first 15 years of the century (Remy et al., 2016). Despite the industry's impressive growth, recent concerns and investigations about its sustainability and impacts on the environment have revealed the destructive consequences of the industrial model. The purpose of this paper is to expand on these ramifications and to develop policies to combat them.

Championing affordability and accessibility in its marketing, the fast-fashion industry has created a new culture of use-and-dispose where the consumer purchases various cheap garments in bulk and disposes of them after a few months of wear. The major companies in the industry, such as H&M, Zara, Uniqlo, Forever 21, and many others, have found major success by implementing a SPA (Specialty Store Retailer of Private Label Apparel) model in which the brands design, produce, and distribute their products in a way that resembles the vertical integration present in various e-commerce platforms like Amazon. The low prices made available by the utilization of SPA has expedited the growth of the use-and-dispose culture, while the endorsement of bulk purchases by online influencers and celebrities along with the prevalent misinformation that donating the clothes after they are no longer wanted is a way to make up for the rampant consumerism have further exacerbated the situation.

The recent mainstream successes of these online influencers, most notably Emma Chamberlain, who has amassed over 27.6 million followers via her 'hauling' videos in which she actively promotes and encourages bulk garment purchases to her young audiences, seem to cement this destructive culture into the youth culture of today. Furthermore, the plethora of online influencers that are either directly sponsored by fast-fashion companies or receive a certain amount of commission for garments sold via their posts in social media makes the use-and-dispose culture even more difficult to get rid of. Even worse, most fashion influencers reside on platforms such as Instagram and Tiktok, which do not enforce a strict policy of disclosing one's sponsorship, rendering their younger audiences even more susceptible to the marketing of the clothing companies. With over "49% of consumers depend[ing] on influencer recommendations" (Digital Marketing Institute, 2021), the use-and-dispose culture that brings these companies record-high sales will die hard.

Fast-fashion and the use-and-dispose culture it fosters possesses massive, global social impacts that must be curtailed. From its direct impact in the form of massive volumes of microplastic and textile waste being dumped into the ocean, to its indirect consequence of disrupting the domestic textile industries of African nations and promoting child labor in low-income nations,

fast-fashion's impacts remain ubiquitous regardless of a nation's prosperity. In this paper, I will explore the three main social impacts of fast-fashion in greater detail and propose two different types of possible policies, supply-sided and demand-sided, to diminish the social impacts of fast fashion.

2. The Social Impacts of Fast-Fashion

As mentioned previously, the main social impacts of fast fashion are threefold: its profitability fuels the full-time child labor industry of Bangladesh and other low-income nations, the discarded garments degrade into microplastic in water and end up harming aquatic life, and the domestic textile industries of African nations are harmed by the influx of used, donated clothing from wealthier nations. In the following subsections, I will elaborate on each social impact's magnitude and corollary effects, which will then lead to a discussion of policies that could possibly curb the social impact of fast-fashion.

2.1. Fast-Fashion and the Expanding Child Labor Industry in Low-Income Nations

The onset of a volume-focused, profit-maximizing fast-fashion industry has resulted in the growth of the child labor industry in many low-income nations, including Vietnam, India, Combodia, Bangladesh, and other nations. Full-time child labor, which does not increase a nation's human capital, perpetuates the cycle of poverty and eliminates the social mobility of families. In countries where child labor is prevalent, many poor families cannot economically sustain themselves with only the income of adults, which pressure them to send their children to work in factories of unsafe working conditions instead of school. This lack of education confines these children to low-paying factory jobs in their adulthood, which again forces their children to work instead of receiving education: the cycle of poverty remains unbroken, with children at constant risk of harm from dangerous equipment and receiving unsustainable income.

Bangladesh is a tangible example of a nation plagued by full-time child labor––"3.5 million [...] children aged from 5-17" have been found to be working in factories, while an estimated 1.2 million children participate in undetectable domestic work (Bangladesh Bureau of Statistics, 2015). 34% of these children are hired by the garment industry in Bangladesh, which makes it the industry that hires the greatest number of child laborers in the nation.

Such a statistic seems to display a discrepancy between corporate policies and reality. Most well-known American and European fast-fashion brands such as Zara and H&M push and advertise their zero child labor tolerance policy and claim that they do not contract with factories that employ children. Yet the statistics indicate that the demand for cheap child labor in Bangladesh by the garment industry has risen, not declined, in the past decade. This is not a case of corporate masking and deception by the fast-fashion giants. Rather, it is a problem of transparency. While bigger clothing companies are legally obligated to be as transparent as possible, smaller fast-fashion companies that supply unbranded clothing to various department stores in America and

European nations are not required to do so. This is reflected in the fashion transparency index of these brands. The index is calculated via the proportion of publicly disclosed information by a certain brand to the total number of categories related to production. For example, if the index assumes 10 different factors in production as its total number of required categories, a brand that discloses 7 of the 10 factors will receive a rating of 70%. The source of the garment and the presence or the absence of child labor in said sources is also a factor in the calculation of the index. While high name-value fashion brands such as H&M, Vans, and The North Face rank among the highest in their transparency (61-70%), the suppliers of unbranded clothing in Costco, Target, and Macy's rank among the lowest (6-10%) (Fashion Transparency Index, 2022). As the latter are not required to reveal the sources of their clothing and materials, sourcing garments from factories that employ child laborers is highly preferable for their low prices. This reveals another layer of the child labor issue: more regulations on popular fashion companies and the source of their garments will not have a sizable impact on reducing child labor in low income nations as most already implement a zero-child labor policy, while economic incentives provided to said factories to not hire children now have a greater risk of backfiring and exploitation.

However, despite the damaging effects of child labor, the complete eradication of the practice may be unachievable in most instances. A second-best, more realistic option to combat it may be allowing child labor but ensuring that children are not forced to full-time labor. Additionally, adequate working conditions and safety, along with a specific required hours in school, if mandated, can reduce the social impacts of child labor in cases where it is necessary. Bangladesh, along with many other low-income nations where child labor remains a great issue, currently does not have such regulations in place. If implemented correctly, families with children that work in factories can be lifted out of the cycle of poverty and the practice itself may be discouraged.

Social mobilization via education is the most effective tool to combat the poverty cycle. The importance of schooling in the increase of human capital simply cannot be understated. In fact, the process has been the main contributor to the reduction of poverty in Brazil. The *Bolsa Familia* was a Brazilian welfare policy enacted in 2003 that sought to promote school participation and investment in children in low-income households by the means of financial incentives. After its enactment, the poverty rate of Brazil saw a continuous decline until 2014, when the Brazilian recession that lasted two years ended the "eleven year streak of poverty reduction" (World Bank, 2020). Although the decrease in the poverty rate was the result of *Bolsa Familia* and many other financial factors, the case of Brazil is a testament to the importance of education in poverty prevention. Other possible solutions to the expanding child labor garment industry of low-income nations will be further elaborated in the supply-side solutions section.

2.2. Fast-Fashion and its Impacts on Aquatic Ecosystems

Plastic pollution in aquatic ecosystems is not caused solely by the garment industry. In fact, it is a problem that is generated by a myriad of other industries that, oftentimes, produce volumes of plastic waste magnitudes greater than that created by the garment industry. From the food industry and the prevalent plastic packaging used to keep vegetables and meats fresh to the technology industry and the microplastics created as a byproduct of hardware production, almost every industry contributes to this problem. It is by no means the purpose of this section to dismiss the gravity of plastic pollution––instead, this section will focus on and address only one aspect of the problem as it is simply too large in scale to elaborate on every source of the pollution.

It is also to be noted that plastic pollution itself is, at part, consumer-driven. It is not solely the result of industrial malice, but rather that of consumer behavior stemming from the convenience and accessibility of plastic. Thus, some parts of plastic pollution can simply be solved by providing the consumers with a financial incentive to not use plastic products or instead to use reusable products. This can be observed in the cases of many nations that have successfully implemented a mandatory plastic bag fee in convenience stores and supermarkets. One tangible example of nations that have found success through financial incentives is South Korea, where a government-mandated 500 KRW (about 0.37 USD) 'environmental fee' on any customers that use a plastic bag was able to dramatically reduce the use of single-use plastic bags. However, as will be elaborated later in this section, most causes of plastic pollution, including garments, cannot be reduced by such simple means.

"By 2050, [the ocean will contain] more plastics than fish by weight" (MacArthur, 2016). In the current status quo, the reality of plastic pollution appears dire. Aside from the sheer mass of plastic in aquatic ecosystems that has led to phenomena like the Great Pacific Garbage Patch three times the size of France, the pollution has also had a devastating effect on the organisms that inhabit the affected areas. The presence of plastic in the ecosystem affects the organisms in two forms: microplastics and macro/mega plastics.

Microplastics, usually classified as plastic that measures between 0.05 - 0.5cm in width (The Ocean Cleanup, 2022), do not directly harm an aquatic organism when directly ingested. Aside from rare cases of direct death via intestinal injuries, microplastics usually build up in an organism's digestive system until it is consumed by a higher predator. Such concentration of microplastics, though initially insignificant, increases exponentially the higher an organism is on the food chain. This process, called biomagnification or bioaccumulation, though lacking a significant effect on organisms located lower the food chain, has a sizable impact on human health. Such an impact is due to the fact that most aquatic organisms consumed by humans, such as tuna, salmon, and other types of fish, are apex predators or secondary carnivores in the respective locations that possess a high concentration of microplastics in their body. Recent research has shown that high levels of microplastic consumption in humans can cause an "alteration in chromosomes which lead to infertility, obesity, and cancer" (Sharma, 2017). Thus, the reduction of microplastic pollution is an integral step in the prevention of plastic pollution-induced diseases in humans.

On the other hand, macroplastics and megaplastics, classified as plastic that measures between 0.5 - 5 cm and 5 - 50 cm in width, respectively (The Ocean Cleanup, 2022), cause direct harm to an aquatic organism via ingestion and strangulation. Aquatic organisms such as sea turtles often mistake plastic for their usual prey, jellyfish, and consume it, which can lead to their death due to choking, internal injury, or starvation. Recently, many seabirds have also starved to death after the ingestion of plastic waste had reduced their stomach's volume. Even in organisms that are able to differentiate plastic from food, megaplastic debris such as packing bands and nets pose a threat of entanglement and strangulation. These entangling materials have been found to affect sea lions and seals greatly, with many sustaining an injury or dying after coming into contact with them (Center for Biological Diversity, 2020). As "92% (of marine debris) interactions are with plastic" and "17% of the species affected by plastic" are endangered (Gall, 2016), the reduction of macroplastic and megaplastic pollution is necessary to maintain the biological diversity in aquatic ecosystems.

How, then, does fast-fashion exactly contribute to the plastic pollution in aquatic ecosystems? The answer is quite simple: garments and textiles are the "largest source of [...] microplastics, accounting for 34.8% of global microplastic pollution" (Somers, 2020). The mechanism in which a garment composed of synthetic materials such as polyester and acrylic releases microfibers, a type of microplastic, is the daily practice of washing them. In one wash, up to 700,000 microfibers are released from the clothing (Napper, 2016), which then directly flow into nearby sources of water. Typically, these fibers cannot be filtered by most filtration systems due to their small size. Though some producers of washing machines do promote a filter that is able to filter out microfibers, they are typically expensive and hard to maintain (Somers, 2020). Even more concerning is the fact that microplastics, once released into a body of water, are nearly impossible to get rid of. Currently, many successful efforts have been made to get rid of macroplastics and megaplastics from oceans and rivers such as Boyan Slat's *The Ocean Cleanup*. However, the current modes of microplastic removal from bodies of water are underdeveloped and unable to filter out meaningful quantities of microplastics. Thus, a more realistic approach to reducing the impact on aquatic ecosystems must revolve around regulations or financial incentives to use an environmentally preferable material that can significantly reduce both microplastic and macroplastic pollution.

2.3. The Use-and-Dispose by Donating Culture and its Impacts on the Textile Industry of African Nations

Perhaps the most prevalent misinformation that followed the spread of the use-and -dispose culture is the notion that donating a piece of clothing after wearing it for a season is beneficial for people in low-income nations, especially the nations in Africa. Though partially true, this assumption omits an important consequence of garment donation––the harming of domestic textile and garment industries in the donee nation. This social impact of fast-fashion engenders a conflict of interest between the consumers and producers of garments in the nation receiving the donated clothing. If this false notion is not corrected, there lies an inherent risk that consumers will develop chronic

justifications for their massive purchases of unsustainable, fast-fashion garments.

In light of this counterintuitive consequence, one may question how the action of donating worn clothing to a charity leads to such consequences in low-income nations. After all, the practice appears to be mutually beneficial, with the donated clothing not ending up in landfills and instead becoming a cheaper alternative for those in low-income nations. To analyze the mechanism in which donated clothing harms the local market, the conflict between consumers and producers, along with the hidden malpractices and misconceptions in garment donation, must be reviewed.

Inspecting the interests of the consumers, it becomes apparent that for these families, especially those of lower incomes, buying cheaper worn clothings in good condition is an opportunity to save money spent on clothing, which accounts for a great proportion of the meager average household income. The consumers of garments in these nations are, indeed, the beneficiaries of the donated clothing. Inspecting the interests of the producers, however, the antithesis holds true: the producers in the garment industry, both the producers of the actual clothing and the materials that compose the clothing, are subsequently harmed, unable to compete with the massive volumes of dirt-cheap garments that flood the market.

More sinister, however, is the hidden business incentives and motives behind the merchants that handle charity-collected garments. While most individuals donate his or her clothing with good intentions and expect their donations to be used altruistically, there exists a lucrative market of selling donated clothing to low-income nations, hidden from the public perception. Referred to as “hidden professionalism”, these donated items fuel an international resale market that promotes the vicious cycle of fast-fashion donation. An example of a “hidden professionalism” market that has seen an explosive growth in recent years is sub-Saharan Africa, “where a third of all globally donated clothes are sold” (Hoskins, 2013). The mechanism behind the market is simple: the large gap between the costs of exporting used clothing to Africa and the profitable resale price has attracted many mercenary merchants attempting to join in on the trade. First, the donated clothes are sold to “second-hand clothing merchants, who sort garments, then bundle them in bales for resale” (Hoskins, 2013). Although exact statistics on the size of this market and the profitability of the practice does not exist, a study has been able to estimate that, for every 300 bales of clothing sold in Africa, the transport costs total about $2,400, while the profits total $30,000 (Brooks et al., 2012).

As long as donations of worn fast-fashion clothing remain profitable, the textile industry of low-income nations will continue to suffer. More awareness must be raised around the fact that donation is not a solution to the fast-fashion problem. Even more importantly, the support of fast-fashion clothing itself must come to an end before any realistic steps can be taken to reduce its social impacts––as of now, the rate of consumption of garments is simply too high for most solutions to actually have an effect.

3. Supply-Sided Solutions to Fast-Fashion's Social Impacts

Supply-sided solutions to the social impacts of fast-fashion are solutions pertaining to changing the behaviors of the producers of fast-fashion garments. In the following subsections, I will be elaborating on two different possible supply-sided solutions, the levying of environmental tax based on garment composition and a cradle-to-grave assessment of garments by fashion brands, that will curtail both the environmental impact of fast-fashion and its impact on domestic textile and garment producers in Africa.

3.1. An Environmental Tax Based on the Composition of Garments

There exists an important dichotomy of materials that are used in the production of a garment: virgin materials and recycled materials. Virgin materials refer to unused raw materials that have not been previously altered or processed, such as raw copper or other metals that have yet been treated (Park, 2007). This includes synthetic materials such as nylons that have just been synthesized from petroleum and woven into fibers for the first time. Recycled materials, on the other hand, refers to a wide range of materials, from materials scrapped directly from a discarded product to materials that have been re-synthesized from its degraded form. The use of recycled materials in producing garments will certainly reduce the environmental burden of resource collection and prevent the exacerbation of plastic pollution, yet the practice remains a rare sight in the garment industry. This absence of recycling materials sourced from used clothing stems from one major factor that controls a garment company's entire business model: cost.

The technology to recycle textiles, shoes, and even ocean plastic has been feasible for a long time. The recycling of textile carpets has become commonplace in the industry, where the used and old carpet is broken down into new fibers and then woven into a carpet of a new design. Similarly, Adidas, since the launch of their collaborative line *Adidas x Parley* in 2015, has been producing shoes composed of recycled ocean plastic to much public accolade and financial success. The technology is also in no way inaccessible to smaller brands––small-scale shoe producers like *Thousand Fell*, too, now offer such a recycled shoe line in their *SuperCircle* project, made available by technological advancements that have made the process affordable to these brands as well.

Despite such access to the recycling technology, garment companies have been reluctant to produce recycled clothing due to the comparatively labor-intensive process of recycling a garment. Unlike shoes, whose main recycled component is a single, homogenous plastic component that is the sole, and carpets, from which the fiber can be extracted with relative ease, garments require a more complex process to recycle, mainly due to the common presence of inner layerings, zippers, and rivets. The widespread use of blended fabrics, such as 98/2 cotton/spandex, also increases the difficulty and cost of garment recycling (Cattermole Consulting Inc., 2019). This renders garment recycling as a costly, slow, and labor-intensive process that is unfavorable to the producers.

For garment recycling to become more common in the status quo, the producers must be given an economic incentive that outweighs the added cost of

recycling. A policy that can achieve this is an environmental tax that is based on the composition of garments. A flexible rate of the environmental tax will be levied onto each garment that is produced by a garment producer, which is determined by the garment composition's proportion of recycled materials to virgin materials. The higher the proportion of recycled materials, the lower the rate of the levied tax, and vice versa. Assuming that it is possible to calculate and implement an optimal rate of this environmental tax on the producers, the added tax burden will motivate the producers to include at least a certain percentage of recycled materials in their garments. It is very important that the tax rate is neither too high that the garment industry is significantly damaged nor too low for it to be dismissable.

Regardless of how many garment producers will actually be initially motivated to incorporate more recycled materials due to the tax, the establishment of garment recycling as a possibility and not an economic *faux pas* also possesses future cultural implications to promote recycling.

3.2. A Cradle-to-Grave Assessment of Garments

As mentioned previously, the influx of donated used clothing to Africa is the main culprit behind the downfall of the domestic textile market in the region. Aside from combating the misinformation that it is good to donate large quantities of used clothing via influencers and public figures not sponsored by fast-fashion companies, which is one aspect of the demand-sided solution in section 4.2, a cradle-to-grave assessment of the garments sold remains as another option to reduce donated clothes.

The term cradle-to-grave refers to the heightened responsibility for the companies that sell garments: they will be accountable for every part of a garment's life cycle, from production to disposal, and any negative social impact that occurs in the process. In this proposed policy, such responsibility will be coordinated with a deposit-refund system, either mandatory or voluntary, that ensures the minimization of disposed or donated clothes. This mechanism can exist both in a voluntary fashion and a mandatory fashion. Even right now, without a regulation that mandates a cradle-to-grave assessment of garments or a deposit-refund system, some fashion companies have voluntarily implemented a deposit-refund system. The *SuperCircle* project, which I have mentioned above as an example of a smaller producer that creates recycled shoes, employs a deposit-refund system. When consumers purchase new sneakers from the project, they are shipped a package that includes the shoes and a prepaid return label, which they can then use to send the shoes back after use. The returned shoes are then recycled and used to produce a new pair of shoes, while the consumer receives a $20 recycling credit that can be used to purchase other items from *SuperCircle*'s affiliate shops. This is an example of clever marketing and an attempt to create a niche for the more environmentally-conscious consumers via a combination of a deposit-refund system and the recycling technology.

However, not all companies may find it profitable or preferable to run such a system, in which case economic incentives such as lowered environmental taxes or direct regulations can be enacted to ensure the implementation of the cradle-to-grave assessment and deposit-refund system.

If a deposit-refund system becomes a widespread practice for most

garment producers and sellers, the amount of donated or disposed clothing will drastically decrease as the consumers of the garments have a greater incentive to return the used clothing instead of simply discarding them in a wastebin or to a charity en masse. If implemented with the environmental tax elaborated on in section 3.1, the clothing companies now also have the incentive to recycle the collected garments and fabrics to produce new garments. Ultimately, this will result in the reduction of plastic pollution caused by discarded garments and the protection of the domestic textile and garment industries in Africa.

4. Demand-Sided Solutions to Fast-Fashion's Social Impacts

Demand-sided solutions to the social impacts of fast-fashion are solutions pertaining to changing the behaviors of the consumers of fast-fashion garments. In the following subsections, I will be elaborating on two different possible demand-sided solutions, the implementation of an intuitive grading system for the environmental impact of a garment and using influencers as a deterrent against fast-fashion, that will discourage consumers from purchasing fast-fashion clothing, thereby effectively curtailing all three social impacts of fast fashion.

4.1. An Intuitive Grading System for Garments

The purchase of fast-fashion clothing remains an issue despite the recent surge in environmental consciousness. Though such a trend is mostly due to the fact that young audiences that are easily swayed are manipulated by fast-fashion-sponsored online influencers, another contributing factor to the issue is the fact that many consumers are unaware of the environmental harm that garment production causes. Even when disregarding the microplastic pollution that is caused by garment production, the garment industry catalyzes environmental harm due to its heavy water consumption. Cotton cultivation itself requires 10,000-20,000 litres of water per kilogram, and an additional 100-150 litres of water is needed to process it into fiber (GLASA, 2015). However, even when provided with such numbers, the impact of a single piece of clothing remains ambiguous to the consumer.

A solution to make the magnitude of the garment's impact more intuitive can actually be borrowed from another sector of conservation––electricity. In South Korea, when purchasing electronics, one would discover that all products are tagged with energy efficiency labels that convert an unintuitive unit for many (kWh) to a more intuitive scale of 1-5. Similarly, garments can be tagged with a label that categorizes each garment from 1-5 depending on the amount of water that was needed to produce the garment, with 1 being the garment that requires the least amount of water to 5 being the garment that requires the most amount of water to produce. This allows the environmentally conscious population to make the conscious decision to either purchase garments with a low environmental impact rating or even fully refrain from purchasing new garments and purchase second-hand clothing. A similar system that encourages the latter is implemented by the second-hand marketplace platform *TheRealReal*, which

provides consumers with an estimate of how much water they saved by purchasing a specific second-hand garment. Aside from water conservation, as the bulk production of fast-fashion garments result in them to require more water than the smaller-scale production of quality garments and to rank higher in the index, the implementation of such a measure will reduce the tendency of the consumers to purchase fast-fashion garments as well.

4.2. Influencers as a Deterrent Against Fast-Fashion

To combat the influence fast-fashion-sponsored influencers have on the easily swayed youth, regulation-based solutions, such as the requirement to fully disclose affiliations and sponsorships on all platforms and not only on YouTube, can be implemented to reduce the efficacy of such influence. However, a more direct approach can also be taken in the form of highly public figures advocating against fast-fashion. The use of public figures to discourage a certain cultural trend has been proven to be effective multiple times.

An instance of such an occurrence is that of Yao Ming and his advocacy against consuming shark fin soup, which was a great social issue at the time due to the decline of the shark population caused by shark fin overconsumption. When the Chinese basketball superstar made a public speech and pleaded to the Chinese population to end shark fin consumption, the shark population was able to make a remarkable recovery following the near-full decline of the yearly shark fin consumption rate. A similar approach can be taken with iconic sports players such as Lebron James, who already owns a premium clothing brand of his own, or musical icons such as Kanye West, who also runs the environmentally conscious clothing line Yeezy, to publicly advocate against fast-fashion and encourage the purchase of quality or second-hand garments in the youth. Although whether these public figure's influence can dominate that of the influencers remains to be seen, if they are able to outweigh the influencers, a rapid decline in the popularity of fast-fashion like that of the shark fin industry may be observed.

5. Discussion

In sections 3 and 4 of the paper, multiple solutions to the social impacts of fast-fashion have been discussed. Though the approach in most of these solutions have been proven to be successful in the context of other issues, whether they will be successful in having the same magnitude of impact on fast-fashion as the original situation remains to be seen. Given the speculative and hypothetical nature of the solutions, further research remains to be conducted to validate the efficacy of the proposed solutions.

References

Cattermole, A. (2019, June 10). *Fiber recycling using mechanical and chemical processes*. Cattermole Consulting Inc. Retrieved August 7, 2022, from https://www.cattermoleconsulting.com/fiber-recycling-using-mechanical-and-chemical-processes/

Digital Marketing Institute. (2018, October 19). *20 Surprising Influencer Marketing Statistics*. Digital Marketing Institute. Retrieved August 11, 2022, from https://digitalmarketinginstitute.com/blog/20-influencer-marketing-statistics-that-will-surprise-you

Gatehouse, J. (2019, July 27). *Will there be more plastic than fish in the ocean by 2050?* CBCnews. Retrieved August 15, 2022, from https://www.cbc.ca/news/politics/ocean-plastic-liberals-fact-check-1.5212632

Hoskins, T. E. (2021, November 18). *Op-Ed: The Trouble with Second-Hand Clothes*. The Business of Fashion. Retrieved August 1, 2022, from https://www.businessoffashion.com/opinions/news-analysis/op-ed-the-trouble-with-second-hand-clothes/

MacArthur, E. (2016, January). *World Economic Forum*

Maxwell, D. (2015). State of the Apparel Sector WATER Report 2. *GLASA*.

Napper, I. and Thompson, R. (2016). Release of synthetic microplastic plastic fibres from domestic washing machines: Effects of fabric type and washing conditions. *Marine Pollution Bulletin*. Retrieved July 21, 2022, from https://www.sciencedirect.com/science/article/pii/S0025326X16307639?via%3Dihub

Ocean Plastics Pollution. Center for Biological Diversity. (n.d.). Retrieved July 30, 2022, from https://www.biologicaldiversity.org/campaigns/ocean_plastics/

Park, C. (2007). A Dictionary of Environment and Conservation. : Oxford University Press. Retrieved 01 Aug. 2022, from https://www.oxfordreference.com/view/10.1093/acref/9780198609957.001.0001/acref-9780198609957.

Rauturier, S. (2022, April 1). *What Is Fast Fashion and Why Is It So Bad?* good on you. Retrieved August 11, 2022, from https://goodonyou.eco/what-is-fast-fashion/

Remy, N., Speelman, E., & Swartz, S. (2020, August 19). *Style that's sustainable: A new fast-fashion formula*. McKinsey & Company. Retrieved August 8, 2022, from https://www.mckinsey.com/business-functions/sustainability/our-insights/style-thats-sustainable-a-new-fast-fashion-formula

Sharma, S., & Chatterjee, S. (2017). Microplastic pollution, a threat to marine ecosystem and human health: a short review. *Environmental science and pollution research international*, *24*(27), 21530–21547. https://doi.org/10.1007/s11356-017-9910-8

Simpliciano, L., Galvin, M., Barry, C., & Williot, D. (2022, July). *Fashion transparency index 2022*. Fashion Revolution. Retrieved August 5, 2022, from https://www.fashionrevolution.org/about/transparency/

Somers, S. (2020, August 25). *Our clothes shed microfibres - here's what we can do...* Fashion Revolution. Retrieved July 10, 2022, from https://www.fashionrevolution.org/our-clothes-shed-microfibres-heres-what-we-can-do/

Theuws, M., Sandjojo, V., & Vogt, E. (2017). Branded Childhood. *SOMO*. Retrieved August 2022, from https://www.stopchildlabour.org/assets/Branded-Childhood.pdf.

The Great Pacific Garbage Patch • The Ocean Cleanup. The Ocean Cleanup. (2022, July 26). Retrieved August 14, 2022, from https://theoceancleanup.com/great-pacific-garbage-patch/

World Bank. (2020). Latin America & the Carribean. *Poverty & Equity Brief*.

Race, Gender, COVID-19, and Oral Health from a Patient and Provider Perspective

Nimrat Kaur

Author Background: *Nimrat Kaur grew up in the United States and currently attends Mercer County Technical Schools – Health Science Academy in Hamilton, New Jersey in the United States. Her Pioneer research concentration was in the field of gender studies/sociology and titled "What Is a Body? Gender, Power, and the Making of a Human."*

Abstract

This paper examines how race and gender have affected the clinical oral health experience within the past two decades through a literature review and demonstrates with original research that race and gender along with the COVID-19 pandemic continue to impact the experience. The original research consists of interviews with seven dental professionals working in pediatric dentistry, an online survey, and pediatric patient observations. These methods helped display the clinical oral health experience from the perspectives of both patients and providers. The results of this study revealed various types of concepts and interesting information, including how communication styles in oral healthcare differ based on gender and race, what parent/guardian accompanies the children to the dentist and how this can differ based on race, and how face masks benefited oral health. Based on the results, it was concluded that race and gender played clear roles in the clinical oral health experience from both perspectives; the pandemic played an ambiguous role.

1. Introduction

Have we ever thought about how our race or gender might influence our oral health experiences? Or how our oral health may be impacted by the COVID-19 pandemic? The gender and race of patients influence how they are treated by dental professionals and the status of their oral health. The decisions of patients in choosing their healthcare provider can impact the experience of the provider and vice versa. The impact or influence these factors have can be positive or negative. In this paper, I argue that race and gender have affected the clinical oral health experience in the United States for the past two decades, and they continue to do so during the pandemic. I investigate how the roles of gender and race have changed with the pandemic for the patient and provider, and how they, along with other variables such as face masks, might have impacted oral health. The literature review below introduces the background and content of this paper with a focus on the past two decades through a thematic organization.

2. Literature Review

2.1. Patient Oral Health Experiences in Terms of Race and Gender in the US

Race and The Patient

In these past two decades, scholars have demonstrated that racial discrimination and disparities exist within oral health care and that race influences the clinical experiences of patients. The groups experiencing discrimination in these medical or dental facilities primarily remain the same as those in other fields. These groups tend to be Black, Hispanic, and Asian. Racial identity and related factors such as socioeconomic status can determine the oral health of many individuals belonging to certain races. For example, untreated cavities are most prevalent in non-Hispanic Black youth and the total number of cavities is most common in Hispanic youth (Fleming and Afful 2018). Poor oral health or oral pain is associated more with certain racial groups such as non-Hispanic Black and non-Hispanic Asian than non-Hispanic white, because of low socioeconomic status and their race (Aldosari et al. 2021). Even if these groups can obtain financial help in accessing dental care, the dental professionals will not always accept their methods. Only one-third of dentists in the US will accept Medicaid because of low reimbursements and patient attitudes (Winegarden and Arduin 2012, 7). Several studies have made evident a link between socioeconomic status and racial disparities in oral healthcare. People (men) belonging to certain races or minorities experience racial disparities due to their low socioeconomic statuses which decrease the chances of them having dental insurance or visiting the dentist to receive proper dental care. This pertains to Black men, who, as a result of this situation, become more prone to greater tooth loss, worse oral cancer survival rates, and more tooth decay in comparison to white men (Lipsky et al. 2021). Race can also impact the clinical oral health experience of a patient through the opinions of dental professionals. A 2018 study conducted in England exhibited how a large majority of dentists had implicit pro-white racial biases, influencing them to recommend extractions to Black patients and root canal treatments to white patients (Patel et al. 2018). Furthermore, patients prefer and have greater satisfaction with care from doctors of the same race (Tanne Hopkins 2002; Penn Medicine 2020). Black patients receive better treatment from Black health professionals, are more engaged with them, and are more likely to consent to preventive services with them (Alsan, Garrick, and Graziani 2019).

Some literature demonstrates the impact of race on oral health with a focus on children and their experiences. Primarily, Black and Hispanic children have always lacked oral healthcare in comparison to their white counterparts. The National Health and Nutrition Examination Survey 2009-2010 shows how Black and Hispanic children had significantly more untreated dental caries and fewer sealants than white children (Dye, Li, and Thornton-Evans 2012). Latino parents, according to a 2008 study, were more likely than other ethnic groups to report their children having poor oral health (Reich et al. 2018). Income is commonly regarded as a factor contributing to racial disparities in healthcare among all ages, including children, and this is valid to a certain extent. Studies have shown how racial disparities have modestly improved throughout the years in low-income Hispanic and Asian children, enabling them to be at the same level of dental care usage as Hispanic white children, but still below those with higher incomes. Yet, non-Hispanic Black children have continued to face

racial disparities in dental care usage at all income levels (Robison, Wei, and Hsia 2020). Why is this the case?

Gender and The Patient

Oral health differences between men and women have been researched in the past two decades and most findings have shown that men have worse oral health than women because of behavioral or biological differences. For example, men are known to engage in risky behaviors and are less likely to engage in preventative care or display positive attitudes regarding dental visits in comparison to women. Even though biologically women are more prone to dental caries, men are more likely to develop root cavities because of dental behaviors such as brushing too hard and not utilizing the recommended fluoride toothpaste (Lipsky et al. 2021). Thus, most previous studies conclude that men have worse oral health than women. Patients perceive the care they receive from their dental providers differently depending on their gender, according to a study conducted in Sudan and published in 2015. The results of this study were, "Patients felt more relaxed when they were being treated by a female dentist, but felt male dentists showed more confidence during the treatment process" (Ibrahim and Awooda 2015). The results of this study are parallel to other similar studies and the stereotypes society has for femininity and masculinity, that females are more empathetic and less career-driven while men are the opposite. However, there is no research showing which perceptions or attitudes providers have towards their patients are influenced by gender, especially in the United States. Very little research in the United States or even globally has ever been done on how parental gender influences the oral health of their children. Research needs to be done to see if oral health differences pertaining to gender have changed with the pandemic. This paper aims to investigate these gender roles during the pandemic or the recent few years.

2.2. Provider Oral Health Experiences in Terms of Race and Gender in the US

Dental professionals also play a role in the clinical oral health experience of the patient and are impacted by race and gender.

Race and The Provider

The profession of dentistry has a large racial and gender gap. Dentistry is known to be a white male-driven field, causing many groups of people to be underrepresented in the dental workforce and an unfair level of hierarchy to be created. Underrepresentation commences in the dental workforce early, in dental school, where the classes are low in diversity (ADEA 2021). White people are more likely to have high-ranking positions and racialized minorities can be bullied and are more likely to be exposed to inequitable disciplinary processes (Jamieson 2021). Furthermore, dental providers of racial minorities treat a disproportionate number of patients with similar racial backgrounds as themselves (Mertz et al. 2016).

Gender and The Provider

Many differences exist between male and female dental professionals, primarily in

treatment and patient preference. Numerous studies show that female dentists treat patients differently than male dentists. In Japan, a study observing the brushing techniques of dental professionals showed that the brushing motion and force differed by gender (Hanasaki et al. 2018). In the United States, female dentists more often advised at-home fluoride treatment than male dentists, who advised treatment in the office (Riley et al. 2011). A 2020 study published in the Journal of Oral Science found that patients who had their oral health evaluated by female dentists were more likely to visit their dentist annually for a dental checkup, in contrast to patients who were checked regularly by male dentists. Additionally, it stated how patients regularly seen by female dentists were more likely to engage in preventive oral health behaviors than patients seen by male dentists (Takeuchi 2020).

Gender in patient preference for providers plays a role in the provider's clinical oral health experience. A 2018 Brazilian study demonstrated the strong preference patients had for female dental professionals (dentists and orthodontists) over males with statistical evidence (Souza-Constantino et al. 2018). An article published in the Journal of Dental Education of the ADEA found in 2004 that anxious patients were more likely to prefer a male dentist than a female dentist, and the percentage varied greatly based on the gender of the patient (Bare and Dundes 2004). In a 2016 literature review done in Chile, a tendency for patients to select same-sex dental professionals was shown to reduce "the shame and fear of physical contact during the exam" (Henríquez-Tejo and Cartes-Velásquez 2016). Not much has been said about how these gender preferences were impacted during the pandemic. There is a large gender gap in the profession of dentistry which affects the clinical oral health experience because it means most patients are being treated by the male gender. The field of dentistry in the United States mainly consists of males, and there is a significant difference in the ratio of male to female dental professionals; in 2021 35.9% of dentists were female and in 2020 approximately 98% of dental hygienists were female (ADA n.d.; ADHA n.d.). Another way gender impacts dental providers is through their salaries, where men typically make a significantly greater amount than women, similar to other career fields (Nguyen Le, Lo Sasso, and Vujicic 2017).

2.3. Pandemic and Oral Health

What has been said about the pandemic and oral health? Has the pandemic worsened oral health and brought factors such as face masks and unemployment to the surface in a new light? With the usage of face masks, people cared less about their dental aesthetics (Suryakumari et al. 2021). Race and gender had a relationship to the frequency of face masks being worn. Black, Asian, and Latino individuals were more likely to wear masks than white individuals. Females were more likely to wear masks more frequently than men (Hearne and Niño 2021). Race and gender also played a role in the unemployment that increased during the pandemic. Hispanic individuals followed by Black individuals lost their jobs the most because of COVID-19 (Acs and Karpman 2020). In regard to gender, women lost their jobs because of COVID-19 more than men (Gezici and Ozay 2020).

Most published literature only relates to the topic of the pandemic and oral health, meaning almost nothing has been said directly about the roles of race and gender during the pandemic in relation to the clinical oral health experience in the United States. The little that is known is that racial disparities have deepened. More research needs to be done on how race and gender have directly impacted oral health during the pandemic.

The Patient and the Pandemic

Patients, including children and adults, were very reluctant to visit the dentist during the pandemic. Almost half of the adults in the United States delayed dental care, which led to negative health consequences (Kranz et al. 2020). Specifically, in children, the pandemic resulted in a decline in their oral health and access to dental care (Lyu and Wehby 2022).

The Provider and the Pandemic

Little to no scholarly literature has been published on how COVID-19 has impacted the business of dentistry along with how race and gender have impacted it. It is known that the salaries of female dentists dropped significantly in comparison with male dentists (Munson et al. 2021). Not much is known about the work-life balance in terms of race and gender during the pandemic in the United States.

3. Methodology

In this paper, mixed methods were utilized to help examine the role of race and gender in the clinical oral health experience. The primary method was interviews; and the others were a survey and patient observations.

Seven dental professionals working at a pediatric dental practice in the state of New Jersey in the United States were interviewed over the course of a month in late spring 2022. Of these seven dental professionals, three were dentists, two dental hygienists, and two dental assistants. One interviewee identifies as a male and he is one of the three dentists. The rest of the dental professionals identify as females. This group consists of racially diverse individuals. Three identify as Asian, two identify as white, one identifies as Black, and one identifies as white Hispanic/Latino. These seven dental professionals are the following:

Dr. O is an employed associate pediatric dentist who identifies as an Asian female and has an education surpassing a graduate degree. Her age is within the range of 25-34 years old.

Cameron is an employed dental hygienist who identifies as a white female with an education consisting of some college. Her age is within the range of 25-34 years old.

Olive is an employed dental hygienist who identifies as a white female with an education consisting of some college. Her age is within the range of 25-34 years old.

Hannah is an employed dental assistant who identifies as a Black or African American female with an education consisting of some college. Her age is within the range of 25-34 years old.

Zaina is an employed dental assistant who identifies as a Hispanic or Latino white female with an education consisting of some college. Her age is within the range of 25-34 years old.

Dr. W is an employed associate pediatric dentist who identifies as an Asian female and has an education consisting of a graduate degree. Her age is within the range of

25-34 years old.

Dr. R is a self-employed pediatric dentist who identifies as an Asian male and has an education consisting of a graduate degree. His age is within the range of 45-54 years old.

I created an online survey using the platform Google Forms and sent out the link to friends, classmates, and peers and asked them to forward the survey to anyone. I posted this survey on a platform called Circle in an academic group consisting mostly of high schoolers and a few adults. Thirty-three individuals from a variety of backgrounds took the survey. Most of them were Asian. 75.8% of these thirty-one individuals identified as female and 24.2% identified as male.

Patient observation studies were conducted over the course of four days for approximately two hours each day. All observations occurred at the two locations of the pediatric dental practice where the interviewees worked. One location accepts Medicaid and the other location accepts PPO, but not Medicaid. Twenty-four patients were observed in total for these four days and all of their dentists were Asian. The race and gender of these patients were noted based on their physical appearance, similarly to the way society infers the race and gender of an individual.

4. Results and Discussion

Race, gender, and the COVID-19 pandemic impact the clinical oral health experience for dental patients and providers. The patients' experience is impacted by dental professionals, health literacy, and insurance (class), as well as the pandemic. The providers' experience is impacted by how they are treated by patients due to race and gender, as well as the pandemic. This section consists of the results and analyses of primarily the interviews I conducted with the seven dental professionals.

4.1. The Patient Experience

Patients and their families at the pediatric office where I interviewed seven dental professionals have had race, gender, and the pandemic play a role in their oral health. I found that the role race, gender, and the pandemic play in the patient experience is not necessarily a discriminatory or negative one.

Provider's Treatment of the Patient

In all seven interviews, it was clear that the tone or attitude of communication from the provider with the patient changes based on their gender. In some interviews, it was demonstrated that the tone or attitude changes based on race.

Cameron is a dental hygienist in the United States who identifies as a white female and her age is within the range of 25-34 years old. I asked Cameron if she thought about her patient's race or gender as a part of providing their care or changed her approach and this was her response:

> Not at all. The one thing that I've noticed, because I don't think about it very much, is that kids are indoctrinated into this gender binary from such a young

> age that like, I'll just call them all “buddy,” because you know, it just means friend, boys and girls can be buddies. And the girls will get mad at me and they'll be like, “I'm not a buddy, I'm a girl.” And I'm like, dude, girls can be buddies, relax. Or sometimes boys, like if I say, “okay, honey, scoot your head up here,” they're like, “I'm a boy. I'm not honey.”

This quote above makes me wonder why we do not think about race? Is it because we associate the word “race” with “racism” and change its meaning to a discriminatory one instead of a physical characteristic? Also, the quote shows the role gender plays in the experience of a dental patient. In society, we have certain norms for genders consisting of names or particular words, and it is clear that these norms follow people into the dental chair. These norms do have an impact on the patient's clinical oral health experience. If you look at the above quote, the words or attitude the dental provider used were perceived negatively by the patients and upset them because of the genders society has associated with the words, “buddy” and “honey.”

Cameron also mentioned that “women tend to be a bit softer and gentler with their patients.”

Hannah, a dental assistant in the United States who identifies as a Black female and is within the 25-34 age group, in her interview mentioned something very similar to Cameron. She said:

> I think I'm a little gentler with the girls. I think with the boys I'm like, you know, and I'm still soft, but I give them like a tough all around. Like, “you got this bud, you got this” but with the girls I'm like, “it's okay” and I think I kind of baby the girls a little more.

Olive is a dental hygienist in the United States who identifies as a white female and her age is within the range of 25-34 years old. She said:

> No, not at all. I think everybody's the same. I treat every patient as if I were to think that they were my kids. And I want them to have an enjoyable experience at the dentist and want to come back.

These three quotes exhibit how female dental providers treat their patients in a more maternal manner. The interviewees used words such as “soft” and “gentle” to describe their attitudes when treating patients and these maternal patterns for female providers were seen across the majority of the interviews. Hannah's response revealed again the gender norms or stereotypes society associates with boys and girls. We believe boys are tougher in nature and need to be treated as such, and girls are sensitive, thus needing gentleness or softness. Therefore, the nature of women to be the more maternal gender and the societal gender norms existing within the clinical oral health experience of patients are clear.

These societal gender norms were also portrayed in the response of Dr. R, an Asian male pediatric dentist and practice owner. Depending on the gender, he changes his tone or word choice, such as, “How you doing, sweetheart?” or, “What's going on, man?”

Dr. O, an associate pediatric dentist who is an Asian female, thinks that

subconsciously she changes her approach based on the gender of the patient. This is evident in one of her responses:

> So, I think subconsciously we do. But I make a conscious effort to not, because I think that there's more than obviously, you know, there's more than just gender that defines a person and you can't make the assumption that a girl is going to be more receptive to Barbie and a boy's more supportive of dinosaurs. Like, that's not necessarily how it is. So, I try and make an effort to not, but I'm sure I do subconsciously.

It's clear that dental providers change their attitudes or tones based on the gender of the patient, which impacts the clinical oral health experience of the patient in a way that can be negative or positive, depending on the patient's perception. But what about race?

Hannah, a dental assistant identifying as a Black female, demonstrates how race can determine the approach, she as a dental provider takes by stating the following:

> Like I can tell maybe a Black mom, "my girl...there's some things you need to do," as opposed to another mom of another race. And I'll say it a little more gentler because I feel like if I say it the way I say it to somebody else, they may not take it that way. So, the advice doesn't change. I think the tone or the way I handle it with the tone, it's a little different. Because I feel like for me sometimes if I saw maybe a Black assistant or whatever, and she said it to me in a certain way, I will take that like, okay, okay. I can get you. You know? So, I noticed that sometimes when I'm giving the parents advice, it's not the advice that changes, but it's definitely the tone and how I say it.

This demonstrates how tone and attitude in communication between a dental provider and patient can change depending on the race of the patient and provider. It can be inferred that Hannah's own race helps her and the patient feel more comfortable communicating, which is a positive thing. Thus, it is likely possible that the approach a dental provider has for a patient can change based on race through the usage of a certain language or tone.

4.2. Race and Health Literacy and Dental Health

Another theme the interviews revealed was that health literacy and the severity of cases is associated with race. Hannah, a dental assistant who identifies as a Black female, said that dental health is all about being knowledgeable. For example, knowing when you can take your child to the dentist to receive a dental check-up. According to her, the communities that tend to be the least knowledgeable when it comes to dental care are the Black and Indian (Asian) communities.

4.3. Class and Oral Health

Throughout the interviews, there were cases when the interviewee would ascribe poorer oral health not necessarily to race and gender, but to socioeconomic factors or class. The majority of the interviewees work at only two dental clinics; one dental clinic only takes PPO, and one takes Medicaid. Olive, a dental hygienist who is a white

female, has noticed the severity of cases with the COVID-19 pandemic to be worse at the dental clinic taking Medicaid, and she says these severe cases are commonly seen in the Hispanic community. Also, Olive related a lack of knowledge among the parents of Medicaid patients to the severity of their cases. This could indicate that their poorer oral health would be because they were hesitant to visit the dentist during the pandemic. However, according to Dr. O, an Asian female pediatric dentist, the Medicaid or lower socioeconomic status patients were actually quicker to come back during the pandemic to receive dental care than the private insurance patients. Why is this? Is this because the Medicaid patients were less educated and did not believe there were any risks with seeing the dentist during the pandemic, or because for them dental care wasn't costly? On the other hand, one interviewee, Zaina, who is a Hispanic/Latino white dental assistant, said that the number of cavities she saw at the two offices were about the same. These results are similar to past research, but also more research needs to be done to investigate the effect of class on oral health after the climax of the pandemic.

4.4. Pandemic and Oral Health

One may likely assume that there has been a tremendous negative change in the oral health status of individuals due to the pandemic because people were afraid to visit the dentist, a place where they are orally "open." However, the interviews I conducted showed mixed results, some saying that there was a major increase in worsened oral health, and some saying there was a slight or no increase at all. A racial influence or pattern was noted, but not necessarily one for gender.

Dr. W, an Asian female pediatric dentist, told me that the number of cavities or oral health is "definitely a lot worse." And Hannah, a dental assistant who is a Black female, agrees with this, because from what she has seen, the severity of cases has changed tremendously for the worse.

Dr. R, a pediatric dentist who identifies as a male, demonstrated the above pattern by saying:

> Races? Yes. Not gender. Of course, you know, people coming from the low-income area who are not well-educated...most commonly are Hispanics and Black people who come from underserved areas or low-income areas. The cavity levels are definitely way higher.

The responses of these dental professionals are not surprising because they are what one would assume to be the case and show patterns parallel to those noted in some past literature. What is surprising are the outlying responses of Zaina and Cameron. For Zaina, a dental assistant who is a Hispanic/Latino white female, and Cameron, a dental hygienist who is a white female, the pandemic has had a very insignificant or no impact on dental caries, and race has had no connection to cavity prevalence.

4.5. The Provider Experience

The interviewed dental professionals have had race, gender, and the pandemic play a role in their careers. This is exhibited in the way they are treated by patients and through the role gender has had in their careers.

Patients' Treatment of the Provider

Zaina, a dental assistant who is a Hispanic/Latino white female, feels that she has been talked down to by the male parents as though she is beneath them, and Hannah, who is a Black female dental assistant, thinks people assume that she isn't very knowledgeable when they view her as a Black woman, when in reality, she knows more than they assume. She also thinks patients ask women more questions to confirm they are giving them the right advice. Furthermore, one female dental provider actually had a stalker at her old job, and believes that men are just more comfortable with having females in their personal space. Gender preference for a dental provider is evidently another way the dental provider experience is impacted.

Cameron says adult patients and, surprisingly, children have a preference for a male or female doctor most of the time, and according to her, the preference among children is shaped by which parent they spend more time with. For instance, there was a Caucasian boy who was sobbing inconsolably while getting his teeth brushed, and said, "I don't want mommy, I want the doctor!" This exemplifies that the boy doesn't see any female as a doctor and signifies a gender disparity. This further means that women are seen to not be as capable as dental professionals in comparison to men and this clearly has a negative impact on their experience as dental providers. Zaina, a dental assistant who is a Hispanic/Latino white female, agrees with this, saying that some pediatric patients prefer a "daddy dentist" over a "mommy dentist," and she believes this is a comfort thing for the patients. The girls don't care about the gender of the provider, and the boys mostly want a male, which is similar to what Cameron said. Cameron said that male patients usually prefer a male dentist and female patients prefer a female dentist.

4.6. Gender in relation to the Patient's Accompaniment

I asked all the interviewees which parent or guardian they see their patient accompanied by the most with a connection to race, and I noticed that the majority of them had very similar answers to this question. It was obvious that the moms or female guardians are the ones accompanying their child more frequently to the dentist, and that in the white population, the mothers/female guardians come often more often, while the fathers/male guardians come more often in the Hispanic and Asian communities. In the Black community, an increase in the number of fathers/male guardians accompanying their children to the dentist was noticed. The reasons why certain parents/guardians of particular races accompany their children more frequently to the dentist are not necessarily because the other parent/guardian is neglecting the children. For instance, colored women may be seen less frequently accompanying their children to the dentist because they have lower salaries than white women, something that many scholarly studies focusing on the wage gap between races have shown. This difference in salary could mean that colored women cannot afford to take off from work due to financial reasons, inflexible schedules, or complex variables at work such as higher-ups with a gender bias, making them unavailable to accompany their children to the dentist.

4.7. The Overall Role of the Pandemic

The pandemic has positively and negatively affected the clinical oral health experience. During the initial stages of the COVID-19 pandemic, when there was great uncertainty regarding the virus, many industries including dentistry struggled

financially. I asked the interviewees about how they believe the COVID-19 pandemic impacted their dental careers and I saw differences based on gender. For example, Dr. R mentioned the shortage of income he as a male practice owner experienced, which portrayed the societal norm of males being the breadwinners of their family, and Zaina, a female dental assistant, discussed how she was put on-call at the dental clinic she used to work at when it was not busy. She mentioned how not having a stable schedule became a problem for her as a mother, making childcare difficult. Zaina's response demonstrates the societal gender norm that mothers are usually the ones concerned for their children.

The literature review of this paper mentions how face masks caused people to be less concerned with their dental aesthetics, and surprisingly, no interviewee believed face masks had a significant negative impact on oral health (Suryakumari et al. 2021). Dr. R, an Asian male who is a pediatric dentist and practice owner, said that face masks are not an excuse to not maintain oral hygiene. Meanwhile, Olive said that parents of patients actually told her face masks benefited their child's oral health by helping them not suck their thumbs, which aided in preventing occlusion abnormalities. These results mean that more research needs to be done on face masks and oral health, because masks are not solely negative for oral health but beneficial in some ways.

4.8. The Survey and Patient Observations

The interview responses had themes parallel to those mentioned in the literature review, some completely different, and some very similar to the results of the survey and patient observation studies I conducted.

The results of the survey exhibited that oral health has not been significantly impacted by the pandemic. Only 3 people out of 33 believed that their oral health had worsened as a result of the pandemic, and 14 of the 33 believed their oral health hygiene habits changed as a result of the pandemic. 11 people believed their oral health hygiene habits increased, 5 people believed they decreased, and 19 people believed they stayed the same. The majority (26 people) said they did not worry less about their dental aesthetics with the usage of face masks. The survey results made it evident that perhaps the connection news or scholarly articles demonstrated between poor oral health and the pandemic or face masks may not be very accurate after all. However, the survey sample size and demographics/locations of the survey participants can potentially limit the accuracy of the conclusion that the pandemic had little effect on oral health.

The results of the patient observation studies at the pediatric dental practice showed that the mothers accompanied their children to the dentist more often than the fathers. In the white community, the mothers could be seen accompanying their children more frequently, and unfortunately, the data I collected for the fathers in terms of race was inadequate, because during the four days I conducted the observations, very few fathers came in. Therefore, this exhibits how mothers are typically the parent more invested in the oral health of their child. Although I was unable to draw conclusions regarding the fathers, for two months I did notice racialized patterns concerning them throughout my internship at the dental practice where I conducted the studies. Those patterns are that in the Hispanic followed by the Asian and Black communities, the fathers accompany their children to the dentist far more often than fathers of the white community. Additionally, the patient observation studies conducted at the practice that does not take Medicaid did not have a single Black patient come in. This may seem very surprising, but it actually is

not considering the location of the practice and ethnic majorities of Medicaid. Essentially, the results of the patient observation studies were very similar to the interview results.

5. Conclusion

The overall finding of this paper is that gender, race, and the pandemic do still impact the clinical oral health experience for both patients and providers. This is important, because our oral health experiences are altered depending on our demographics. Racial disparities still exist within oral healthcare and within the same groups as mentioned in previous literature. Gender continues to play a crucial role in the way the patient and provider are approached and treated. Although some effects of race and gender remain the same as prior to the pandemic, some have changed. The pandemic has an ambiguous role in the clinical oral health experience, since it has had a mix of negative and positive effects on oral health. All of this information was drawn from seven interviews, a survey, and patient observation studies that each had their own limitations due to demographics or scale. I plan to further explore the pandemic and its relation to oral health and why society associates thinking of physical characteristics with negative actions. I hope to discover the clinical oral health experience to be more equal across all genders and races.

Bibliography

Acs, Gregory, and Michael Karpman. “Employment, Income, and Unemployment Insurance during the COVID-19 Pandemic,” June 2020. https://www.urban.org/sites/default/files/publication/102485/employment-income-and-unemployment-insurance-during-the-covid-19-pandemic.pdf.

ADEA. “‘Very Little Progress’ in U.S. Dental Schools Enrolling Black Students.” JDE article: "Very little progress" in U.S. dental schools enrolling black students, April 20, 2021. https://www.adea.org/Press/Apr2021-JDE-dental-school-black-enrollment/.

Aldosari, Muath, Suellen da Mendes, Ahad Aldosari, Abdullah Aldosari, and Mauro Henrique de Abreu. “Factors Associated with Oral Pain and Oral Health-Related Productivity Loss in the USA, National Health and Nutrition Examination Surveys (NHANES), 2015–2018.” *PLOS ONE* 16, no. 10 (2021). https://doi.org/10.1371/journal.pone.0258268.

Alsan, Marcella, Owen Garrick, and Grant C. Graziani. “Does Diversity Matter for Health? Experimental Evidence from Oakland - NBER,” August 2019. https://www.nber.org/system/files/working_papers/w24787/w24787.pdf.

Bare, Lyndsay C., and Lauren Dundes. “Strategies for Combating Dental Anxiety.” *Journal of Dental Education* 68, no. 11 (2004): 1172–77. https://doi.org/10. 1002/j.0022-0337.2004.68.11.tb03862.x.

“Dentist Workforce.” American Dental Association. Accessed June 26, 2022. https://www.ada.org/resources/research/health-policy-institute/dentist-workforce.

Dye, Bruce A., Xianfen Li, and Gina Thornton-Evans. “Oral Health Disparities as Determined by Selected Healthy People 2020 Oral Health Objectives for the United States, 2009–2010,” August 2012. https://permanent.fdlp.gov/ gpo44730/db104.pdf.

Fleming, Eleanor, and Joseph Afful. “Prevalence of Total and Untreated Dental Caries among Youth: United States, 2015-16,” July 25, 2018. https://www.cdc.gov/nchs/data/databriefs/db307.pdf.https://www.cdc.gov/nchs/data/databriefs/db307.pdf

Gezici, Armagan, and Ozge Ozay. “How Race and Gender Shape Covid-19 Unemployment Probability.” SSRN, August 17, 2020. https://ssrn.com/ abstract=3675022.

Hanasaki, Mika, Kuniko Nakakura-Ohshima, Tsutomu Nakajima, Yukiko Nogami, and Haruaki Hayasaki. “Gender Difference of Tooth Brushing Motion and Force on Self-Brushing and Caregivers' Brushing in Dental Professionals.” *Dental, Oral and Craniofacial Research* 4, no. 4 (2018). https://doi. org/10.15761/docr.1000258.

Hearne, Brittany N., and Michael D. Niño. “Understanding How Race, Ethnicity, and Gender Shape Mask-Wearing Adherence during the COVID-19 Pandemic: Evidence from the COVID Impact Survey - Journal of Racial and Ethnic Health Disparities.” SpringerLink. Springer International Publishing, January 19, 2021. https://link.springer.com/article/10.1007/s40615-020-00941-1.

Henríquez-Tejo, Rocío Belén, and Ricardo Andrés Cartes-Velásquez. “Patients’ Perceptions about Dentists a Literature Review,” May 2016. http://www.scielo.edu.uy/pdf/ode/v18n27/en_v18n27a03.pdf.

Hopkins Tanne, Janice. “Patients Are More Satisfied with Care from Doctors of Same Race.” BMJ: British Medical Journal. BMJ, November 9, 2002. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1124573/.

Munson, Bradley, Rachel Morrissey, Brittany Harrison, and Marko Vujicic. “How Did Covid-19 Affect Dentist Earnings.” American Dental Association, September 2021. https://www.ada.org/resources/research/health-policy-institute/dental-practice-research/how-did-covid-19-affect-dentist-earnings.

Ibrahim, Haifaa Mohamed, and Elhadi Mohieldin Awooda. “Comparison of Patients Perception of Dental Care Offered by Male or Female Dentist: Cross-Sectional Hospital Based Study.” *European Journal of General Dentistry* 4, no. 03 (2015): 117–20. https://doi.org/10.4103/2278-9626.163329.

Kranz, A.M., G. Gahlon, A.W. Dick, and B.D. Stein. “Characteristics of US Adults Delaying Dental Care Due to the Covid-19 Pandemic.” *JDR Clinical & Translational Research* 6, no. 1 (2020): 8–14. https://doi.org/10.1177/ 2380084420962778.

Jamieson, L. “Racism and Oral Health Inequities; an Introduction.” Community dental health. U.S. National Library of Medicine, May 28, 2021. https://pubmed.ncbi.nlm.nih.gov/33848410/.

Lipsky, Martin S., Sharon Su, Carlos J. Crespo, and Man Hung. “Men and Oral Health: A Review of Sex and Gender Differences.” *American Journal of Men's Health* 15, no. 3 (2021): 155798832110163. https://doi.org/10.1177/ 15579883211016361.

Lyu, Wei, and George L. Wehby. “Effects of the COVID-19 Pandemic on Children’s Oral Health and Oral Health Care Use.” *The Journal of the American Dental Association*, 2022. https://doi.org/10.1016/j.adaj.2022.02.008.

Mertz, Elizabeth, Jean Calvo, Cynthia Wides, and Paul Gates. “The Black Dentist Workforce in the United States.” *Journal of Public Health Dentistry* 77, no. 2 (2016): 136–47. https://doi.org/10.1111/jphd.12187.

Nguyen Le, Thanh An, Anthony T. Lo Sasso, and Marko Vujicic. “Trends in the Earnings Gender Gap among Dentists, Physicians, and Lawyers.” *The Journal of the American Dental Association* 148, no. 4 (2017). https://doi.org/ 10.1016/j.adaj.2017.01.005.

“Oral Health Fast Facts & Stats.” Accessed June 26, 2022. https://www.adha.org/resources-docs/72210_Oral_Health_Fast_Facts_&_Stats.pdf.

Patel, N., S. Patel, E. Cotti, G. Bardini, and F. Mannocci. “Unconscious Racial Bias May Affect Dentists’ Clinical Decisions on Tooth Restorability: A Randomized Clinical Trial.” *JDR Clinical & Translational Research* 4, no. 1 (2018): 19–28. https://doi.org/10.1177/2380084418812886.

Reich, Stephanie M., Kristin S. Hoeft, Guadalupe Díaz, Wendy Ochoa, and Amy Gaona.

"Disparities in the Quality of Pediatric Dental Care: New Research and Needed Changes." *Social Policy Report* 31, no. 4 (2018): 1–27. https://doi. org/10.1002/ sop2.2.

Riley, Joseph L., Valeria V. Gordan, Kathleen M. Rouisse, Jocelyn McClelland, and Gregg H. Gilbert. "Differences in Male and Female Dentists' Practice Patterns Regarding Diagnosis and Treatment of Dental Caries." *The Journal of the American Dental Association* 142, no. 4 (2011): 429–40. https://doi.org/10. 14219/jada.archive. 2011.0199.

Robison, Valerie, Liang Wei, and Jason Hsia. "Racial/Ethnic Disparities among US Children and Adolescents in Use of Dental Care." *Preventing Chronic Disease* 17 (2020). https://doi.org/10.5888/pcd17.190352.

Souza-Constantino, Andréa Maria, Ana Cláudia de Castro Ferreira Conti, Leopoldino Capelloza Filho, Sara Nader Marta, and Renata Rodrigues de Almeida-Pedrin. "Patients' Preferences Regarding Age, Sex, and Attire of Orthodontists." *American Journal of Orthodontics and Dentofacial Orthopedics* 154, no. 6 (2018). https://doi.org/10.1016/j.ajodo.2018.02.013.

"Study Finds Patients Prefer Doctors Who Share Their Same Race/Ethnicity." Pennmedicine.org, November 9, 2020. https://www.pennmedicine.org/news/ news-releases/2020/november/study-finds-patients-prefer-doctors-who-share-their-same-race-ethnicity.

Suryakumari, Achanta, Sangeetha Sasidharan, Dhatri Majji, and Divya Uppala. "'Mask Mouth' During COVID - 19 Pandemic -A Myth or A Truth," 2021. https://www. researchgate.net/profile/UppalaDivya/publication/352366291_Mask_Mouth_During_COVID_-_19_Pandemic_A_Myth_or_A_Truth/links/60c61fcea6fdcc2e613e27d5/Mask-Mouth-During-COVID-19-Pandemic-A-Myth-or-A-Truth.pdf.

Takeuchi, Kenji, Yuki Noguchi, Yukie Nakai, Toshiyuki Ojima, and Yoshihisa Yamashita. "Dentist Gender-Related Differences in Patients' Oral Health Behaviour." *Journal of Oral Science* 62, no. 1 (2020): 32–35. https://doi.org/10.2334/josnusd.18-0462.

Winegarden, Wayne, and Donna Arduin. "The Benefits Created by Dental Service Organizations," October 2012. https://www.pacificresearch.org/wp-content/ uploads/2017/06/DSOFinal.pdf.

Patient Centered Medicine: Evolution of the FDA's Drug Approval Regulations The Thalidomide Tragedy in 1961 and the AIDS Crisis in the 1980s

Chujun Liu

Author Background: *Chujun Liu grew up in China and currently attends Grier School in Tyrone, Pennsylvania in the United States. Her Pioneer research concentration was in the field of history/sociology and titled "In Sickness and in Health: Topics in the History and Sociology of Public Health."*

Abstract

The United States Food and Drug Administration (FDA) made significant revisions to its drug approval regulations after two serious public health disasters, the thalidomide tragedy in 1961 and the HIV/AIDS crisis in the 1980s. After the thalidomide tragedy, the FDA strengthened its drug approval regulations and gained authority to monitor the entire new drug innovation process. However, after facing the AIDS crisis in the 1980s where AIDS patients had no effective drug available, the FDA reduced the amount of testing required to quickly approve drugs for life-threatening diseases. The FDA continued to refine its drug approval regulations in response to health threats and diseases and gradually became more patient-centered and transparent.

1. Introduction

Over the last 100 years, the United States Food and Drug Administration (FDA)[1] has continued to refine its drug approval regulations to better protect the public in response to several large public health disasters. In order to understand the pattern of its changes, this paper will examine two extremely influential public health events in the United States: the thalidomide tragedy in 1961 and the Acquired Immune Deficiency Syndrome (AIDS) crisis in the 1980s. The thalidomide tragedy pushed the FDA to adjust its inadequate drug approval protocols and design stricter drug review policies, whereas the AIDS crisis in the 1980s led the FDA to reconsider its inflexible regulations and establish nuanced regulations for different illnesses. The FDA also gradually became more patient-

[1] The U.S. Food and Drug Administration, founded in 1906, oversees public health by ensuring the safety and efficacy of medicines, biological products, and medical devices. The FDA constantly adjusts its protocols in responding to public health crises.

centered and transparent during this revision process. This paper will discuss the factors that led to such differences by analyzing what happened during these two public health disasters, exploring their root cause, and comparing the fundamental differences regarding the ethics of drug usage in these two events.

The thalidomide tragedy was a devastating adverse drug event that resulted from the use of thalidomide, a drug prescribed to treat nausea. It caused more than 10,000 birth defects around the world (mostly in Germany), before being recalled in 1961.[2] The United States avoided such a disaster since thalidomide was not approved in the U.S. However, this tragedy triggered an internal review of the U.S.'s drug approval procedures. Congress passed the more rigorous 1962 Drug Amendment, which clearly defined the required materials for new drug applications.[3] In the 1980s, the emergence of AIDS, a life-threatening disease for which there were no effective drugs, led to a large number of deaths in the U.S.[4] This crisis led to a significant change to the FDA's drug approval regulations by triggering the FDA to develop more nuanced regulations for different types of diseases.[5] For example, for life-threatening diseases with no available treatment, the FDA eliminated the long-term effectiveness data requirements, to ensure that patients could receive drugs sooner.[6]

Throughout history, there have been a wide range of debates over drug ethics, and drug approval is just one aspect of this. Scholars have debated issues such as the requirement of informed consent, the conflict of interest among different groups during drug development, the use of placebos in clinical trials for life-threatening diseases, and the scope of consideration regarding public health that determines drug regulations.[7] A 1962 editorial, "Guinea Pigs and

---

[2] Thalidomide was a sedative that was also used to treat other symptoms, like insomnia, and almost no patient complained that the drug had any side effects. Arthur Daemmrich, "A Tale of Two Experts: Thalidomide and Political Engagement in the United States and West Germany," *Social History of Medicine* 15, no. 1 (2002): 138, http://rave.ohiolink.edu/ejournals/article/329581259.

[3] The 1962 Drug Amendment allowed the FDA to monitor the complete drug development process from animal testing to clinical trials and set a standard process for new drug innovation. For a detailed discussion, see: "U.S. Food and Drug Administration, *Summary of the Drug Amendments of 1962* (Washington, DC: Government Printing Office, 1962), 2-5, https://hdl.handle.net/2027/hvd.32044032092983. Daniel Carpenter, "Reputation and Power Crystallized: Thalidomide, Francis Kelsey, and Phased Experiment, 1961-1966," In *Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA* (New Jersey: Princeton University Press, 2010), 282, https://ebookcentral.proquest.com/lib/oberlin/detail.action?docID=540270.

[4] Centers for Disease Control and Prevention, "Current Trends Mortality Attributable to HIV Infection/AIDS -- United States, 1981-1990," Morbidity and Mortality Weekly Report, accessed June 22, 2022, https://www.cdc.gov/mmwr/preview/mmwrhtml/00001880.htm.

[5] David Vogel, "AIDS and the Politics of Drug Lag," *Public Interest*, Summer 1989, 7-8, https://www.proquest.com/docview/1298111856?accountid=12933&imgSeq=1.

[6] William H. Eaglstein, "Brief History of the FDA," in *The FDA for Doctors* (Cham: Springer International Publishing, 2014), 92, http://rave.ohiolink.edu/ebooks/ebc/9783319083629.

[7] A placebo is a harmless pill or procedure with no therapeutic value that is prescribed for the psychological benefit of patients and for forming a control group in clinical trials. Informed consent is the permission for practices that a patient provides to a doctor that

People," argued that patients who participate in drug testing without being informed of their situation were being deprived of their human rights.[8] Thirty years later, Charles McCarthy's argument of waiving informed consent under emergencies adds more complexity to this topic. "In some circumstances," noted McCarthy, "informed consent should be waived to allow emergency research to go forward" as demanding consent from severely injured patients may delay the optimal treatment time.[9]

Conflict of interest existing among different parties was also discussed by many scholars. Howard Brody, a bioethicist and family physician, contends that during drug development, pharmaceutical companies, doctors, and patients have complex and sometimes competing interests, and this sometimes hinders the drug development process.[10] Researchers Lisa Cosgrove, Sheldon Krimsky, Emily Wheeler, Shannon Peters, Madeline Brodt, and Allen Shaughnessy also assert that in order to provide patients the best treatment, physicians should follow a guideline that is "free from conflicts of interest."[11]

Debates over the use of placebos have emerged with regard to fatal diseases. John Porter, Bruce Forrest, and Ann Kennedy argue that it is unethical to use placebos in clinical trials for AIDS and other life-threatening diseases. Even though using placebos does provide more credible results, it is unethical not to do everything possible to save lives.[12] Sara Sorscher, Azza AbuDagga, Sammy Almashat, Michael Carome, and Sydney Wolfe found through research that it is still very common to provide only placebos instead of currently available treatments to control groups in clinical trials of life-threatening diseases. Nevertheless, they argue that such practices pose potential risks to human subjects and the FDA should seriously assess if the benefit of using placebos outweighs the possible adverse impact it has on control groups.[13]

Still, some scholars focus on whether the FDA's drug regulation is broad

---

demonstrated their acknowledgment of the potential risks and benefits.

[8] "Guinea Pigs and People," *The Christian Century* 79, no. 33 (August 15, 1962): 975, http://ezproxy.oberlin.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=a6h&AN=ATLA0000672848&site=ehost-live&scope=site.

[9] Charles R. McCarthy, "To Be or Not to Be: Waiving Informed Consent in Emergency Research," *Kennedy Institute of Ethics Journal* 5, no. 2 (June 1995): 156-7, https://muse.jhu.edu/article/245759.

[10] Howard Brody, "The Ethics of Drug Development and Promotion: The Need for a Wider View," *Medical Care* 50, no. 11 (November 2012): 910, https://www.jstor.org/stable/41714598.

[11] Lisa Cosgrove et al., "Conflict of Interest Policies and Industry Relationships of Guideline Development Group Members: A Cross-Sectional Study of Clinical Practice Guidelines for Depression," *Accountability in Research: Policies and Quality Assurance* 24, no. 2 (2017): 110, http://ezproxy.oberlin.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=120156574&site=ehost-live&scope=site.

[12] John D.H. Porter, Bruce D. Forrest, and Ann R. Kennedy, "The Ethics of Placebos in AIDS Drug Trials," *HEC Forum* 4, no. 3 (May 1992): 157-9, http://rave.ohiolink.edu/ejournals/article/329207529.

[13] Sarah Sorscher et al., "Placebo-only-controlled versus Active-controlled Trials of New Drugs for Nine Common Life-threatening Diseases," *Open Access Journal of Clinical Trials* 10 (January 2018): 26, https://doaj.org/article/a9ee5c240239417da39df35bbd733d35.

enough to achieve public health. Food and Drug law experts Patricia J. Zettler, Margaret Foster Riley, and Aaron S. Kesselheim argue that FDA's drug regulations often only "narrowly focused on weighing the benefits and risks regarding the product itself," while for drugs that have externalities, like opioids, the FDA should also take a more "public health" perspective when evaluating its approval.[14]

Scholars come at the issues of drug ethics from different points of view, but debates over the thalidomide tragedy and the AIDS crisis in the 1980s are most central to the discussion of catalysts for changing the drug approval process. Many scholars have presented opinions toward the drug approval regulations before and after the thalidomide tragedy by comparing the strictness and effect of those policies. Others also shared their perspectives on drug lag during the AIDS crisis.

Pharmaceutical scientist Arthur Daemmrich, Economist Mary Olson, FDA specialist Peter Barton Hutt, and Historian Robert Temple all argue that the U.S. drug review system prior to the thalidomide tragedy was not very stringent and failed to protect the public from unsafe drugs.[15] Daemmrich believed that the United States avoided the disaster solely because Frances Kelsey, the FDA reviewer for thalidomide, paid close attention to data submitted about this drug and was extra careful about the review process.[16] According to Ellen Rice, "only the right person, in the right place, at the right time had saved them from tragedy."[17] All scholars who have commented on this topic essentially agree that the U.S. drug approval regulations prior to the thalidomide tragedy were not strict enough to protect the safety of the population. The U.S. Congress also was aware of this issue and revised some FDA policies.[18]

In 1962, Congress unanimously passed the 1962 Drug Amendment; however, scholars present different attitudes toward the Amendment. The 1962 editorial "Guinea Pigs and People," for example, advocated for the new Amendment as it finally required informed consent before clinical trials.[19] Forty-six years later, Suzanne White Junod, a historian in the FDA History Office, also praised the 1962 Drug Amendment for the same reason.[20] Mary Olson celebrated

---

[14] Patricia J. Zettler, Margaret Foster Riley, and Aaron S. Kesselheim, "Implementing a Public Health Perspective in FDA Drug Regulation," *Food and Drug Law Journal* 73, no. 2 (2018): 221, https://www.jstor.org/stable/26661176.

[15] Daemmrich, "A Tale," 153; Mary K. Olson, "The Food and Drug Administration (1962-Present)," in *Guide to U.S. Health and Health Care Policy*, ed. Thomas R. Oliver (California: SAGE Publications, 2014), 68, https://ebookcentral.proquest.com/lib/oberlin/detail.action?docID=1810523; Peter Barton Hutt and Robert Temple, "Commemorating the 50th Anniversary of the Drug Amendments of 1962," *Food and Drug Law Journal* 68, no. 4 (2013): 451, https://heinonline.org/HOL/P?h=hein.journals/foodlj68&i=491.

[16] Daemmrich, "A Tale," 153.

[17] Ellen Rice, "Dr. Frances Kelsey: Turning the Thalidomide Tragedy into Food and Drug Administration Reform" (Senior Division Research Paper), 5, https://studylib.net/doc/8769719/.

[18] U.S. Food and Drug Administration, *Summary of the Drug Amendments of 1962*, 2-5.

[19] "Guinea Pigs," 975.

[20] Suzanne White Junod, "FDA and Clinical Drug Trials: A Short History," *U.S. Food and Drug Administration*, 3, accessed May 25, 2022, https://www.fda.gov/media/110437/download.

the Amendment due to another reason. She noted that after the Amendment "the FDA was no longer a helpless bystander" because the FDA finally had the power to monitor pharmaceutical companies during new drug development processes.[21] On the other hand, Public Health professor David Dranove, Economist Sam Peltzman, and Richard E. Faust (a drug company researcher) all criticized the 1962 Drug Amendment, arguing that it caused drug lag and increased costs.[22]

Criticism of drug lag reached its peak during the AIDS crisis in the 1980s. In 1980, James Scheuer (1920-2005), a Democratic member of the United States House of Representatives openly charged that "the FDA is contributing to needless suffering and death of thousands because it is denying them life-saving and life-enhancing drugs that are available abroad far sooner than they are here."[23] The risk level was lower for drugs treating AIDS. In a life-or-death situation, any drugs that could have extended patients' life were beneficial and should have been approved quickly. However, the 1962 Drug Amendment failed to accomplish this. David Vogel contends that after the 1962 Drug Amendment, new drugs took approximately ten years to reach the market, and people with AIDS could not wait such a long time.[24] In response to this situation, AIDS activists advocated for changes to the drug approval process, which pushed the FDA to pass faster approval regulations for life-threatening diseases without effective treatment.[25] Jessica Pace, Narcyz Ghinea, Ian Kerridge, and Wendy Lipworth noted in their article that the new regulation for life-threatening diseases that lacked drugs was a significant change to the strict 1962 Drug Amendment and was more moral since it saved a lot of AIDS patients.[26] However, drugs passed through accelerated approval were not always effective. Indeed, Chul Kim, MD and Vinay Prasad, MD contend that 86% of the accelerated approval drugs between 2008 and 2012 "fail[ed] to show gains in survival" and had adverse side effects.[27]

Many scholars have praised the FDA for making drug approval

---

[21] Olson, "The Food," 68.

[22] Drug lag is a situation in which a new drug already receives approval and is marketed in other countries but is not available in the U.S. David Dranove, "The Costs of Compliance with the 1962 FDA Amendments," *Journal of Health Economics* 10, no. 2 (July 1991): 235-6, https://www.sciencedirect.com/science/article/pii/0167629691900069?via%3Dihub; Sam Peltzman, "An Evaluation of Consumer Protection Legislation: The 1962 Drug Amendments," *Journal of Political Economy* 81, no. 5 (1973): 1087, http://www.jstor.com/stable/1830639; Richard E. Faust, "The Impact of the 1962 Drug Amendments on the Research Process," *Managerial and Decision Economics* 1, no. 4 (1980): 201, http://www.jstor.org/stable/2487329.

[23] James Scheuer, quoted in Vogel, "AIDS and the Policies," 75.

[24] Vogel, "AIDS and the Policies," 74.

[25] Robert W. Hansen, Paul L. Ranelli, and L. Douglas Ried, "Stigma, Conflict, and the Approval of Aids Drugs," *Journal of Drug Issues* 25, no. 1 (January 1995): 133, http://rave.ohiolink.edu/ejournals/article/345386989.

[26] Jessica Pace et al., "Accelerated Access to Medicines: An Ethical Analysis," *Therapeutic Innovation and Regulatory Science* 51, no. 2 (March 2017): 159, http://rave.ohiolink.edu/ejournals/article/347180047.

[27] Chul Kim and Vinay Prasad, "Cancer Drugs Approved on the Basis of a Surrogate End Point and Subsequent Overall Survival: An Analysis of 5 Years of US Food and Drug Administration Approvals," *JAMA Internal Medicine* 175, no. 12 (December 2015): 1993, https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2463590.

regulations more stringent after the thalidomide tragedy, but some also criticized that the strict policy led to drug lag, which hindered AIDS patients from receiving treatments. The FDA's response to the two disasters was very different even though both crises posed a great danger to public safety, and this paper will discuss the factors that led to such differences.

2. The Thalidomide Tragedy

To place the thalidomide tragedy in context, it is important to first understand the history of U.S. drug approval regulations. Before the thalidomide tragedy, the United States followed the 1938 Food, Drug, and Cosmetic Act, which only required pharmaceutical companies to submit a New Drug Application (NDA) before marketing. Materials needed for the submission of the NDA include information about the chemistry of the drug, toxicity test results done by the drug company, and any clinical trial data and supplementary materials from the company.[28] However, this regulation was not very stringent. New drugs under this Act received automatic approval for marketing sixty days after their NDA submission unless the FDA identified problems.[29] Drugs did not need to be formally "approved" by the FDA to enter the market. Under this policy, if the FDA did not identify a potential problem in a drug within sixty days, they missed the chance of preventing the dangerous drug's initial entrance to the market. Once a drug was on the market, it was costly to recall it. Besides, multitudes would have already purchased this dangerous drug before the drug was completely recalled from every drug store.

In addition, because pharmaceutical companies were only required to submit new drug applications after all preparatory work had been completed rather than submitting their designed methods prior to actual operations, the FDA was unable to oversee clinical trial processes. There were also no standard testing protocols required, so pharmaceutical companies could test the drug by any means without notifying the FDA.[30] This meant that even if the clinical trial was designed poorly or if the drug harmed patients, the FDA would not get notified immediately, which put the safety of the public at risk. Finally, the 1938 Food, Drug, and Cosmetic Act did not require the practice of informed consent. Hence, some physicians did not tell their patients that the drug they prescribed was still in the testing phase.[31] The 1962 Editorial "Guinea Pig and Peoples" angrily stated the situation the U.S. patients were in: "human beings, informed or not, willing or not, are the guinea pigs for drug concerns and the doctor."[32] Looking from a twenty-first-century perspective, the 1938 Food, Drug, and Cosmetic Act was far too lenient and could not protect the public's safety. Yet few people at that time questioned such regulations since the regulations prior to this were even less stringent.[33] People's perspectives, however, on drug safety, shifted

[28] Junod, "FDA and Clinical Drug," 5.
[29] Olson, "The Food," 66.
[30] Junod, "FDA and Clinical Drug," 6.
[31] "Inside Story of a Medical Tragedy: Exclusive Interview with Dr. Frances O. Kelsey," *U.S. News and World Report*, 1962, 54, https://search-ebscohost-com.ezproxy.bowdoin.edu/login.aspx?direct=true&db=rel&AN=89256256&site=ehost-live.
[32] "Guinea Pigs," 975.
[33] In 1958, Senator Estes Kefauver (D-Tennessee) argued for a more stringent drug

significantly after the thalidomide tragedy.

Between 1957 and 1961, doctors in Germany prescribed thalidomide to treat morning sickness during pregnancy, and the drug was extremely effective. Thalidomide was also commonly used as a sedative, treating problems such as insomnia and tension.[34] Moreover, physicians found that patients could not use thalidomide to commit suicide as they might with other sleeping pills.[35] For instance, there were reports about adults taking six times more than the usual amount without any adverse reaction.[36] Thalidomide was thought not to have any serious side effects (because birth defects did not emerge immediately due to the pregnancy period) and such characteristics made it a popular drug in Germany. However, in 1961, many cases of phocomelia (a form of birth defect of which children are born with missing limbs) were reported by physicians related to the use of thalidomide. Before being recalled in late 1961, thalidomide caused more than 10,000 birth defects around the world.[37]

Thalidomide did not enter the U.S. drug market thanks to the work of FDA reviewer Frances Oldham Kelsey. Although the United States did not have stringent drug approval regulations, Kelsey paid extremely careful attention to the drug company's application, preventing a similar tragedy in the U.S. After first reviewing the material submitted by Merrell, the pharmaceutical company that wanted to bring thalidomide to market in the U.S., Kelsey concluded that the data provided was insufficient to prove that the drug was safe. As Kelsey mentioned in an interview, Merrell did not submit any data indicating that the long-term use of thalidomide was safe.[38] Since thalidomide had excellent performance in Germany and such data was not mandatory according to regulations, this was not a critical problem. Yet, to stay on the safer side, Kelsey withheld the application and requested more data.[39] As Kelsey waited for more information, reports on birth defects in Germany emerged in the U.S. media. After the terrible defects appeared in the U.S. newspapers, Merrell quietly withdrew the drug application.[40] The United States escaped the disaster thanks to the good work of Frances Kelsey.

However, the United States was still significantly influenced by this tragedy. The FDA had no right to oversee drug companies during clinical trials at

---

approval regulation. However, his proposal received little support at the time. It was not until after the thalidomide tragedy that people revisited Kafauver's hearings and seriously considered its feasibility. Congress revised his hearings in 1962 and turned it into the 1962 Drug Amendment. For a detailed discussion, see Junod, "FDA and Clinical Drug," 9.

[34] Daemmrich, "A Tale," 138-9.

[35] "Inside Story," 54.

[36] Daemmrich, "A Tale," 138.

[37] Dranove, "The Costs," 235; Daemmrich, "A Tale," 138.

[38] Frances Oldham Kelsey, "Autobiographical Reflections," interview, U.S. Food and Drug Administration, 63, accessed May 31, 2022, https://www.fda.gov/media/89162/download.

[39] FDA clinical reviewers had the right to hold a drug application for 60 days (not approving or rejecting it). After deciding to hold a drug, FDA reviewers had to write a thorough letter to the pharmaceutical company on the problems they identified with the drug, and the pharmaceutical company should submit supplementary material regarding the identified issue. However, the FDA reviewers at the time often chose not to hold an application, because it took a lot of effort to write that thorough letter to the drug company; Kelsey, "Autobiographical Reflections," 63-4.

[40] Carpenter, "Reputation and Power," 266.

the time. Therefore, Merrell tested thalidomide in the U.S., which led to a total of ten phocomelia cases.[41] Numerous media outlets in the U.S. also published reports about the tragedy in Germany, so almost everyone heard about the drug at the time. Many people who saw the tragedy in Germany expressed hope for more stringent drug review policies. A poll from *Washington Star* in 1962 showed that 76.3% of the respondents wanted stricter control over drugs.[42] Moreover, during the thalidomide tragedy, Kelsey discovered that Merrell was irresponsible when sending out samples of thalidomide for testing. As Kelsey noted in her interview, "They (physicians) were told that the drug was virtually ready to be approved and, in essence, it was a detailing procedure to get them familiar with this drug."[43] The drug was sent out casually since Merrell was so confident about its safety and efficiency, but such practice obviously had safety issues. If Merrell was irresponsible regarding testing, it was likely that many other pharmaceutical companies also did not pay much attention to the accuracy of data collection during clinical trials. More importantly, the regulations at the time did not explicitly state that such practice was illegal, which posed a huge risk to the public. The Merrell case and the thalidomide tragedy revealed shortcomings of the U.S. drug approval regulations. It proved that the FDA's drug review regulations failed to protect the safety of the public. Even if Merrell strictly followed every rule, public safety risks remained due to the less stringent regulations at the time. If it was not for Frances Kelsey who paid extra attention when reviewing the application, thalidomide might have led to a similar disaster in the U.S.[44]

As issues emerged after the thalidomide tragedy, it was clear that the U.S. drug approval regulations at the time were highly problematic and had to be revised. As a result, in late 1962, Congress unanimously passed the more rigorous 1962 Drug Amendment. Under this Amendment, drug applications had to be formally approved by the FDA prior to entering the market, and the time limit to review the NDA increased to 180 days.[45] Furthermore, drug companies had to submit animal toxicity test data that showed safety and premarket testing plans (called INDs) to the FDA that described their detailed design for three phases of clinical trials.[46] Informed consent also became mandatory after the amendment.[47] The 1962 Drug Amendment gave the FDA full authority to monitor

[41] Hutt and Temple, "Commemorating the 50th," 451.
[42] Carpenter, "Reputation and Power," 282.
[43] Kelsey, "Autobiographical Reflections," 72.
[44] Daemmrich, "A Tale," 152-3.
[45] Hutt and Temple, "Commemorating the 50th," 452.
[46] IND is shortened for Investigational New Drug. Under the 1962 Drug Amendment, drug companies were required to perform three phases of clinical trials, which was never required in previous regulations. Phase I aims to find out the safety and dosage of the new drug. Phase II tests the efficacy and side effects of the drug. Phase III is the most time-consuming of all three stages, requiring the involvement of from 300 to 3000 volunteers (depending on the disease), and it monitors the efficacy and adverse reactions of the drug. Olson, "The Food," 67; U.S. Food and Drug Administration, "Step 3: Clinical Research," The Drug Development Process, accessed July 24, 2022, https://www.fda.gov/patients/drug-development-process/step-3-clinical-research.
[47] Junod, "FDA and Clinical," 11; The 1962 Drug Amendment also revised other previous regulations. For a detailed discussion, see: "U.S. Food and Drug Administration, *Summary of the Drug Amendments of 1962* (Washington, DC: Government Printing Office, 1962), https://hdl.handle.net/2027/hvd.32044032092983.

the whole drug investigational process since pharmaceutical companies had to notify the FDA of their plans for clinical trials. This ensured that FDA knew of all the clinical trials that were performed in the U.S., so drugs could not enter clinical trials without notifying the FDA.

Compared to the 1938 Food, Drug, and Cosmetic Act, the 1962 Drug Amendment was much stricter and was able to better protect the public's safety. By granting the FDA power to oversee clinical trials, public health risks during clinical trials declined. The three phases of clinical trials provide the FDA with better data on the safety and efficacy of new drugs, which helps them better decide whether the new drug should be approved. Requiring informed consent before any testing also ensured transparency to patients. However, this strict drug review regulation was not completely flawless. Pharmaceutical companies complained about the increased cost when innovating new drugs. According to Richard Faust, the research and development spending for a new drug had more than quadrupled to over $1.3 billion.[48] "New drug R & D (research and development) is viewed as less rewarding," noted Faust, and very few drug companies were willing to develop new drugs after the Amendment was first established.[49] Due to the disinterest of drug companies and the time-consuming new drug development process, there were very few new drug applications during the next decade. Nevertheless, since drugs commonly used by people were available on the market and there were no major medical incidents during this period, the negative impact of the 1962 Drug Amendment was never very significant. However, the AIDS crisis in the 1980s, another significant public health crisis, led to a reconsideration of such policies.

3. The AIDS Crisis in the 1980s

The 1962 Drug Amendment led to drug lag. Because the more stringent drug review policies required drug companies to demonstrate fully the drug's safety and effectiveness and perform three phases of clinical trials, drug companies needed a longer time for testing prior to submitting an NDA. The review process for new drugs also took much longer. The FDA now had 180 days to review the application, and the agency often requested more time.[50] "The average drug took approximately ten years to get through the FDA's testing process," noted Vogel, "four times longer than it took before 1962."[51] Drug lag always existed after the 1962 Drug Amendment; however, the AIDS crisis in the 1980s brought this issue to a peak, as there finally was someone who asked for a change.

In the 1980s, HIV/AIDS emerged as a fatal disease without effective treatment. Between 1981 and 1990, more than 100,000 people died from AIDS.[52]

---

[48] Faust, "The Impact," 201.
[49] Faust, "The Impact," 202.
[50] Louis Lasagna, "Congress, the FDA, and New Drug Development: Before and after 1962," *Perspectives in Biology and Medicine* 32, no. 3 (1989): 335-6. https://muse.jhu.edu/article/402324/pdf.
[51] Vogel, "AIDS and the Policies," 74.
[52] Centers for Disease Control and Prevention, "Current Trends Mortality Attributable to HIV Infection/AIDS -- United States, 1981-1990," Morbidity and Mortality Weekly Report, accessed June 22, 2022, https://www.cdc.gov/mmwr/preview/mmwrhtml/00001880.htm.

Life expectancy for people diagnosed with AIDS ranged from a few months to several years, and the mortality rate was as high as 80% (and possibly even higher because some deaths were not recorded or occurred after this statistic was published).[53] In 1981, the beginning of the AIDS crisis, there was no effective drug on the market. Many pharmaceutical companies started to develop AIDS drugs, but under the 1962 Drug Amendments, the new drug investigational process might take nearly ten years.[54] However, people with AIDS could not wait that long; therefore, AIDS activists started to protest for faster approval of potential drugs. Many of the activists were angry with the FDA, believing that "they were more interested in maintaining the scientific standards of clinical trials than in providing new options for the thousands of patients who were dying as a result of HIV infection."[55] The FDA was on the edge of losing public trust, which pushed them to again adjust their protocols.

As AIDS activists kept demanding policy changes, the FDA issued a series of regulations for severe diseases without effective treatment starting in 1983. In June 1983, the FDA enacted the Treatment IND, which allowed the FDA to "approve a treatment protocol for any patient with a serious disease" during the investigational process. [56] This protocol granted AIDS patients with severe symptoms access to drugs that were still in the clinical trial stages. In 1988, the FDA issued another policy, allowing people with AIDS to import drugs from foreign countries that were not yet approved in the United States.[57] Following this regulation, the FDA also suspended the phase 3 testing requirement for AIDS and other severe diseases that lacked effective treatment by allowing the distribution of drug during phase 3 testing, in an effort to speed up new drug development.[58] Further, in 1992, the FDA passed the Accelerated Approval Regulation, which expedited the NDA review process for drugs that treat life-threatening diseases.[59] These new policies shortened the development time of new drugs, and the FDA

[53] James W. Curran et al., "Epidemiology of HIV Infection and AIDS in the United States," *Science* 239, no. 4840 (February 5, 1988): 610, https://www.jstor.org/stable/1700181.

[54] Vogel, "AIDS and the Policies," 74.

[55] Eve K. Nichols and Institute of Medicine (US) Roundtable for the Development of Drugs and Vaccines Against AIDS, "Historical Perspective," in *Expanding Access to Investigational Therapies for HIV Infection and AIDS: March 12–13, 1990 Conference Summary* (Washington, D.C.: National Academy Press, 1991), 11, https://ebookcentral.proquest.com/lib/oberlin/detail.action?docID=3375988.

[56] Nichols and Institute of Medicine (US) Roundtable for the Development of Drugs and Vaccines Against AIDS, "Historical Perspective," 15.

[57] Before this protocol, and owing to safety concerns, the FDA forbade the importation of foreign drugs that were not approved in the United States. The FDA limited the amount of medicine imported to a 3-month dose to ensure importers did not make money by selling those drugs in the US; Vogel, "AIDS and the Policies," 79.

[58] Usually, there were 3 stages in a clinical trial. Stage 3 requires the drug company to test the drug in a larger population and demonstrate that the drug is effective. However, phase 3 of the clinical trial takes a long time, usually a year or longer, to gather sufficient data; Vogel, "AIDS and the Policies," 80.

[59] Lewis A. Grossman, "AIDS Activists, FDA Regulation, and the Amendment of America's Drug Constitution," *American Journal of Law and Medicine* 42, no. 4 (November 2016): 727, http://rave.ohiolink.edu/ejournals/article/347466180.

approved the first AIDS drugs (AZT, ddl, ddc) within a few years. [60] Thanks to AIDS activists, the first AIDS drug – AZT – only took about 25 months to enter the market.[61] Compared to the 1962 Drug Amendment, these new regulations for AIDS and other life-threatening diseases were much less strict. However, the drug approval policy in the U.S. became more nuanced after this crisis as the FDA started to have different approaches for different illnesses.

HIV/AIDS in the 1980s was a life-threatening disease without effective treatment, so the most likely outcome for people with AIDS was death. However, thalidomide was used to treat nausea, which was a much simpler symptom that could be treated with other medicines. Therefore, when facing two very different situations, the FDA needed different approaches. John McKie and Jeff Richardson discussed an idea called "the rule of rescue" in 2003, which is "an imperative people feel to rescue identifiable individuals facing avoidable death."[62] This concept could be applied to the AIDS crisis in the 1980s. Even though AIDS was not an avoidable death, some drugs could extend patients' life. Most people with AIDS face death in a few years after diagnosis. Given the circumstances, any drug that might be effective was worth trying, as it might help patients lengthen their life expectancy. Therefore, although the drug approval regulation was not very strict for AIDS after the changes, the policies ensured AIDS patients experimental treatments as soon as possible. "In the context of an inevitably fatal disease," noted Lewis A. Grossman, "victims might not demand the same level of certainty as people suffering from less serious ailments."[63] The U.S. Court of Appeals for the Tenth Circuit also questioned, "[W]hat can . . . 'safe' and 'effective' mean as to such persons who are so fatally stricken with a disease for which there is no known cure?"[64] Many agreed that in such life-or-death situations, the risk of approving a potentially beneficial drug was much lower since most patients were more willing to accept some minor adverse effects if the drug could help them live longer.

However, the situation regarding thalidomide was very different. The thalidomide tragedy happened at the FDA's early stage, during a time when the FDA had negligible say in setting the standards for drug approval. Furthermore, thalidomide was not the only drug available to treat nausea, and it led to serious birth defects. The cost of using thalidomide was much higher than the benefit. Under this circumstance, the use of thalidomide could and should be avoided. On the contrary, AIDS, as a life-threatening illness, could accept higher risks owing to the potential of AIDS drugs to prolong patients' lifespans. There must exist a balance between effectiveness and risk. In the scales of AIDS, these two aspects were relatively balanced. However, the scales of the thalidomide tragedy tipped toward risk more than effectiveness. Treating nausea should not require as tragic a cost as phocomelia. Therefore, to protect the public's safety and well-being,

---

[60] Grossman, "AIDS Activists," 728-9.

[61] Lucas Richert, "Reagan, Regulation, and the FDA: The US Food and Drug Administration's Response to HIV/AIDS, 1980-90," *Canadian Journal of History* 44, no. 3 (2009): 471-2, https://web.p.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=0&sid=92e6c9a0-b746-44e1-86b0-ef86f65a920b%40redis.

[62] John McKie and Jeff Richardson, "The Rule of Rescue," *Social Science and Medicine* 56, no. 12 (June 2003): 2407, http://rave.ohiolink.edu/ejournals/article/334811599.

[63] Grossman, "AIDS Activists," 697.

[64] Rutherford v. United States, 582 F.2d 1234, 1237 (10th Cir. 1978).

stringent regulations were necessary for drugs such as thalidomide that treat non-life-threatening illnesses, but not for drugs treating AIDS and other life-threatening diseases.

4. Conclusion

Drug regulations continue to evolve today as more public health crises happen that push the FDA to rethink its policies. According to Dr. Yajie Li, the FDA has recently introduced stricter requirements for the use of surrogate endpoints.[65] Surrogate endpoints are indicators used by pharmaceutical companies that indirectly reflect the results of clinical trials and help predict whether the drug is potentially effective. They are often used in chronic diseases since the results of long-term effectiveness, such as life expectancies, are not available until much later. For example, researchers use a patient's HIV viral load before and after drug treatment to determine whether the drug would be effective in treating AIDS. Surrogate endpoints can help pharmaceutical companies to predict the effectiveness of drugs, but "the FDA did not have a very clear understanding of how the surrogate endpoints worked a few years ago," noted Dr. Li in 2022. "Therefore, the FDA approved several surrogate endpoints that couldn't actually reflect the effectiveness of the drug."[66] However, Dr. Li noted that the FDA now restricts the use of surrogate endpoints – only those that have been validated for accuracy in multiple experiments may be used.[67] This policy ensures that the drugs being approved using surrogate endpoints are truly effective.

The FDA's ultimate purpose is to protect public safety. In order to do this, the FDA needs to ensure the safety of drugs. However, in certain cases like the AIDS crisis in the 1980s, public health and safety can be better maintained by not hindering patients from accessing potentially beneficial drugs. For minor illnesses, strict regulations for drugs are needed to protect the public, but when faced with life-threatening diseases like AIDS, it is part of the FDA's responsibility to address patient needs as quickly as possible. The FDA can only protect the public's health and safety to the greatest extent possible by designing drug approval regulations that are responsive or reflect specific situations.

As the FDA continues to revise its drug approval regulations, it also has become more patient-centered and transparent. Public safety is its primary concern, and the FDA has been adjusting to best protect the public and maximize health. The thalidomide tragedy and the AIDS crisis in the 1980s are two very clear-cut cases, but the FDA has also found solutions when dealing with less explicit situations. For example, the FDA has always closely monitored the drug instructions submitted by pharmaceutical companies. Drug instructions must always clearly state the potential adverse side effects one could have after using the drug. Patients could decide whether or not to use the drug according to their situation. There is never a one-way solution when dealing with drugs and public health, and the FDA should be flexible to change in order to best protect the

[65] Yajie Li (MD) is the vice president at Parexel China and a former clinical reviewer at the National Medical Products Administration's Center of Drug Evaluation; Yajie Li, interview by the author, Beijing, China, June 27, 2022.
[66] Li, interview by the author.
[67] Li, interview by the author.

public. If drugs are approved too quickly without careful inspection, another thalidomide tragedy might occur. Yet, if drugs are approved too slowly, then there might be another AIDS crisis. Therefore, the FDA should make judgments based on the specifics of the disease and the availability of drugs. By informing patients of the potential risks they might face when taking medicines or participating in trials, the FDA could also better protect public health and safety.

Bibliography

Brody, Howard. "The Ethics of Drug Development and Promotion: The Need for a Wider View." *Medical Care* 50, no. 11 (November 2012): 910-12. https://www.jstor.org/stable/41714598.

Carpenter, Daniel. "Reputation and Power Crystallized: Thalidomide, Francis Kelsey, and Phased Experiment, 1961-1966." In *Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA*, 257-327. New Jersey: Princeton University Press, 2010. https://ebookcentral.proquest.com/lib/oberlin/detail.action?docID=540270.

Centers for Disease Control and Prevention. "Current Trends Mortality Attributable to HIV Infection/AIDS -- United States, 1981-1990." Morbidity and Mortality Weekly Report. Accessed June 22, 2022. https://www.cdc.gov/mmwr/preview/mmwrhtml/00001880.htm.

Cosgrove, Lisa, Sheldon Krimsky, Emily E. Wheeler, Shannon M. Peters, Madeline Brodt, and Allen F. Shaughnessy. "Conflict of Interest Policies and Industry Relationships of Guideline Development Group Members: A Cross-Sectional Study of Clinical Practice Guidelines for Depression." *Accountability in Research: Policies and Quality Assurance* 24, no. 2 (2017): 99-115. http://ezproxy.oberlin.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=120156574&site=ehost-live&scope=site.

Curran, James W., Harold W. Jaffe, Ann M. Hardy, W. Meade Morgan, Richard M. Selik, and Timothy J. Dondero. "Epidemiology of HIV Infection and AIDS in the United States." *Science* 239, no. 4840 (February 5, 1988): 610-16. https://www.jstor.org/stable/1700181.

Daemmrich, Arthur. "A Tale of Two Experts: Thalidomide and Political Engagement in the United States and West Germany." *Social History of Medicine* 15, no. 1 (April 2002): 137-58. http://rave.ohiolink.edu/ejournals/article/329581259.

Dranove, David. "The Costs of Compliance with the 1962 FDA Amendments." *Journal of Health Economics* 10, no. 2 (July 1991): 235-38. https://www.sciencedirect.com/science/article/pii/016762969190069?via%3Dihub.

Eaglstein, William H. "Brief History of the FDA." In *The FDA for Doctors*, 89-93. Cham: Springer International Publishing, 2014. http://rave.ohiolink.edu/ebooks/ebc/9783319083629.

Faust, Richard E. "The Impact of the 1962 Drug Amendments on the Research Process." *Managerial and Decision Economics* 1, no. 4 (1980): 201–3. http://www.jstor.org/stable/2487329.

Grossman, Lewis A. "AIDS Activists, FDA Regulation, and the Amendment of America's Drug Constitution." *American Journal of Law and Medicine* 42, no. 4 (November 2016): 687-742. http://rave.ohiolink.edu/ejournals/article/347466180.

"Guinea Pigs and People." *The Christian Century* 79, no. 33 (August 15, 1962): 975-76. http://ezproxy.oberlin.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=a6h&AN=ATLA0000672848&site=ehost-live&scope=site.

Hansen, Robert W., Paul L. Ranelli, and L. Douglas Ried. "Stigma, Conflict, and the Approval of Aids Drugs." *Journal of Drug Issues* 25, no. 1 (January 1995): 129-39. http://rave.ohiolink.edu/ejournals/article/345386989.

Hutt, Peter Barton, and Robert Temple. "Commemorating the 50th Anniversary of the Drug Amendments of 1962." *Food and Drug Law Journal* 68, no. 4 (January 2013): 449. https://advance.lexis.com/api/document?collection=analytical-materials&id=urn:contentItem:5B6R-H830-00CV-V0F1-00000-00&context=1516831.

"Inside Story of a Medical Tragedy: Exclusive Interview with Dr. Frances O. Kelsey." *U.S. News and World Report*, 1962, 54-55. https://search-ebscohost-com.ezproxy.bowdoin.edu/login.aspx?direct=true&db=rel&AN=89256256&site=ehost-live.

Junod, Suzanne White. "FDA and Clinical Drug Trials: A Short History." In *A Quick Guide to Clinical Trials*. Excerpt from *A Quick Guide to Clinical Trials*. Accessed May 25, 2022. https://www.fda.gov/media/ 110437/download.

Kelsey, Frances Oldham. "Autobiographical Reflections." Interview. U.S. Food and Drug Administration. Accessed May 31, 2022. https://www.fda.gov/media/89162/download.

Kim, Chul, and Vinay Prasad. "Cancer Drugs Approved on the Basis of a Surrogate End Point and Subsequent Overall Survival: An Analysis of 5 Years of US Food and Drug Administration Approvals." *JAMA Internal Medicine* 175, no. 12 (December 2015): 1992-94. https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2463590.

Lasagna, Louis. "Congress, the FDA, and New Drug Development: Before and after 1962." *Perspectives in Biology and Medicine* 32, no. 3 (Spring 1989): 322-43. https://muse.jhu.edu/article/402324/pdf.

Li, Yajie. Interview by the author. Beijing, China. June 27, 2022.

McCarthy, Charles R. "To Be or Not to Be: Waiving Informed Consent in Emergency Research." *Kennedy Institute of Ethics Journal* 5, no. 2 (June 1995): 155-62. https://muse.jhu.edu/article/245759.

McKie, John, and Jeff Richardson. "The Rule of Rescue." *Social Science and Medicine* 56, no. 12 (June 2003): 2407-19. http://rave.ohiolink.edu/ejournals/article/334811599.

Nichols, Eve K., and Institute of Medicine (US) Roundtable for the Development of Drugs and Vaccines Against AIDS. "Historical Perspective." In *Expanding Access to Investigational Therapies for HIV Infection and AIDS: March 12–13, 1990 Conference Summary*, 5-18. Washington, D.C.: National Academy Press, 1991. https://ebookcentral.proquest.com/lib/oberlin/detail.action?docID=3375988.

Olson, Mary K. "The Food and Drug Administration (1962-Present)." In *Guide to U.S. Health and Health Care Policy*, edited by Thomas R. Oliver, 65-78. California: SAGE Publications, 2014. https://ebookcentral.proquest.com/lib/oberlin/detail.action?docID=1810523.

Pace, Jessica, Narcyz Ghinea, Ian Kerridge, and Wendy Lipworth. "Accelerated Access to Medicines: An Ethical Analysis." *Therapeutic Innovation and Regulatory Science* 51, no. 2 (March 2017): 157-63. http://rave.ohiolink.edu/ejournals/article/347180047.

Peltzman, Sam. "An Evaluation of Consumer Protection Legislation: The 1962 Drug Amendments." *Journal of Political Economy* 81, no. 5 (September/October 1973): 1049-91. http://www.jstor.com/stable/1830639.

Porter, John D.H., Bruce D. Forrest, and Ann R. Kennedy. "The Ethics of Placebos in AIDS Drug Trials." HEC Forum 4, no. 3 (May 1992): 155-62. http://rave.ohiolink.edu/ejournals/article/329207529.

Richert, Lucas. "Reagan, Regulation, and the FDA: The US Food and Drug Administration's Response to HIV/AIDS, 1980-90." *Canadian Journal of History* 44, no. 3 (2009): 467-88. https://web.p.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=0&sid=92e6c9a0-b746-44e1-86b0-ef86f65a920b%40redis.

Sorscher, Sarah, Azza AbuDagga, Sammy Almashat, Michael A. Carome, and Sydney M. Wolfe. "Placebo-only-controlled versus Active-controlled Trials of New Drugs for Nine Common Life-threatening Diseases." *Open Access Journal of Clinical Trials* 10 (January 2018): 19-28. https://doaj.org/article/a9ee5c240239417da39df35bbd733d35.

U.S. Food and Drug Administration. *Summary of the Drug Amendments of 1962*. Washington, DC: Government Printing Office, 1962.https://hdl.handle.net/2027/hvd.32044032092983.

U.S. Food and Drug Administration. "Step 3: Clinical Research." The Drug Development Process. Accessed July 24, 2022. https://www.fda.gov/patients/drug-development-process/step-3-clinical-research.

Vogel, David. "AIDS and the Politics of Drug Lag." *Public Interest*, Summer 1989, 73-85. https://www.proquest.com/docview/1298111856?accountid= 12933&imgSeq=1.

Zettler, Patricia J., Margaret Foster Riley, and Aaron S. Kesselheim. "Implementing a Public Health Perspective in FDA Drug Regulation." Food and Drug Law Journal 73, no. 2 (2018): 221-56. https://www.jstor.org/stable/26661176.

Using Sentiment Analysis, Statistical Analysis, and Neural Network Simulations to Analyze and Simulate the Correlation Between Cyberspace Freedom and Development

Jason Zhuang

Author Background: *Jason grew up in China and currently attends Shenzhen College of International Education in Shenzhen, Guangdong, China. His Pioneer research concentration was in international relations/STS and titled "Understanding Global Cyber Power."*

Abstract

In this paper, I explored the correlation between cyberspace freedom and development on the scale of individuals, states, and the system of world politics. I found that there is a significant positive correlation between cyberspace freedom and countries' long-term development potential as well as political stability. Also, as artificial intelligence has become more mature and has illustrated its ability to aid in simulations of real-world issues like global warming and violent conflicts, I harnessed the strength of neural networks to simulate the whole system of international politics and how internet freedom influences a country's development and survival. Last but not least, with the sentiment analysis toolkit Natural Language Toolkit (NLTK), I also examined the change in online attitudes towards a country when measures restricting cyber-freedom are implemented, as well as long-term trends in countries' cyberspace reputations.

1. Introduction

Cyberspace connectivity is one of the main drivers behind modern societal developments. By providing connections for people across the globe, making management information systems possible in real life (Cyberspace 75), and enabling academic cooperation on an unimaginable scale, cyberspace has dramatically benefited humanity.

In this research paper, I will examine cyberspace freedom's influence on subjects from the individual level to the state level, then to the level of the abstract system of freedom and development.

At the individual level, I examined cyberspace freedom's effect on teenagers' academic performance. The correlation between the two variables suggests that students having more cyberspace freedom positively correlates with academic performance in the internationally standardized test PISA.

At the state level, I examined cyberspace freedom's effect on countries' Human Development Index, Gross Domestic Product, and political stability. When states are discontent with the freedom individuals can obtain in cyberspace, they implement measures to restrict citizens' usage and access to cyberspace. Authoritarian countries fear a second "Arab Spring" event and are doubling down on censorship and disconnection from the global internet. Splinternets, regional networks that are not part of the global network, are emerging around the globe, not just in Russia, China, and Iran *(Mainwaring)*. Although creating intranets and Splinternets and limiting cyberspace freedom, in general, seems like foolproof plans for states, there are hidden consequences to limiting people's cyberspace freedom. After researching, I found that cyberspace freedom directly affects economic development, human development, political stability, and academic performance. I also collected online discussion data from popular discussion website Reddit to analyze the effects of increasing censorship on states' online reputations in four case studies involving Russia, China, Thailand, and New Zealand.

At the system level, I modeled different simulations of world politics and analyzed strategies and decisions made by the neural network. The neural networks simulate strategies that help countries survive in the game of world politics, and the results from the neural network could educate and inspire world leaders on the impacts of cyberspace freedom.

2. Related Work

As previous studies by Perez and Ben-David have shown, internet/cyberspace freedom cannot be narrowly defined by its immediate results--such as contributing to immediate livelihood or any other narrow objects. The researchers collected data from internet usage of more than two thousand computers for one year (Perez and Ben-David, chap.3.2). After parsing the dataset and analyzing the websites most visited by people in Airjadi network, researchers concluded that internet enhancd freedom in all of Sen's freedom categories. This research, while worthwhile, restricted its research to a specific area, which is less representative of the whole system of cyberspace freedom and development.

Law Professor Foram Patel expertly proved that a proper approach to providing cyberspace freedom and ensuring the security of people online is to leave legislation to the private sector. There will be no "one size fits all" solution for regulations and controls in cyberspace. Also, one's responsibility for cyberspace security should be shouldered by one alone. Dr. Patel further states that cyberspace/web offers an extraordinary stage for self-articulation. It advances law-based qualities and allows us to communicate and impart our perspectives and insights to others (Patel, Foram). My research tested the claim by Professor Foram and his conclusion that states' cyberspace restrictions will yield negative results for states.

3. Correlation Between Internet Freedom and Development

3.1 The Data Collection

- The first dataset was the *Freedom house internet freedom index in selected countries in 2021*. The dataset includes 64 countries and their assigned internet

freedom index. The index is judged on three criteria: A. Obstacles to Access; B. Limits on Content; C. Violations of User Rights. *("Freedom House Index")*

- The second database is the Political stability index *("Political Stability by Country, around the World")*. The dataset includes two columns of variables—country name and political stability index. This data is obtained from media articles, protest records and political regime change databases.
- World 2020 Gross Domestic product database is from the world bank *(GDP (Current US$) | data)*. There are two columns of data—country name and gross domestic product value in dollars.
- Database from FactsMaps of 2018 PISA scores of students worldwide *(FactsMaps)*.
- Database of the top 20 Countries with the highest rate of cybercrime includes data about cybercrime rates in each country. Each country lists six contributing factors, the share of malicious computer activity, malicious code rank, spam zombies rank, phishing website hosts rank, bot rank, and attack origin to substantiate its cybercrime ranking *(Sumo3000)*.

3.2 Method of Examining the Correlations Between Data

To examine the correlation between variables and to obtain an objective correlation coefficient, I will use Pearson's Correlation Coefficient, namely

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

The value of r indicates the strength of a correlation; the closer the absolute value of r is to 1, the stronger the existing correlation. A positive r indicates a positive correlation, while a negative r indicates a negative correlation.

3.3 The Correlation Between Internet Access and Teenager Academic Performance

Teenagers usually have the best knowledge about cyberspace, from being fluent in various online gaming software, frequently engaging in online discussions, and following the latest edition of "dank memes" (an urban diction invented by the youth to describe fascinating pieces of viral infographics). However, teenagers are also the ones that suffer the most from cyberspace restrictions, from parental control to laws restricting teenagers' cyberspace access.

With the increasing prevalence of improper internet usage and addictions to cyber content in teenagers, many parents are wary of their children's use of cyberspace and are limiting their cyberspace usage. A study has shown that problematic gamers negatively influence family cohesion and damage their future *(Bonnaire and Phan)*. Monitoring a child's internet use can effectively reduce the chance of addiction, improper use, and mental harm from allowing children to roam in cyberspace *(Hill)*.

However, some parents may be over-restricting their children's internet usage. Overbearing in cyberspace can cause damage to a child's development potential by reducing their creativity and restricting a crucial source of information.

Similarly, some governments are putting direct limits on teenagers' online usage—from directly banning specific age groups from accessing websites to limiting cyberspace activity time to after school and during weekends. These restrictions encroach on teenagers' cyberspace freedom, even though governments claim it is for addiction reduction *(Buckley)*.

PISA is the OECD's Programme for International Student Assessment. PISA measures 15-year-olds' ability to use their reading, mathematics, and science knowledge and skills to meet real-life challenges *(PISA)*. This particular dataset is chosen to measure the academic abilities of teenagers in different countries due to its cultural fairness and holistic evaluation of teenagers' academic abilities.

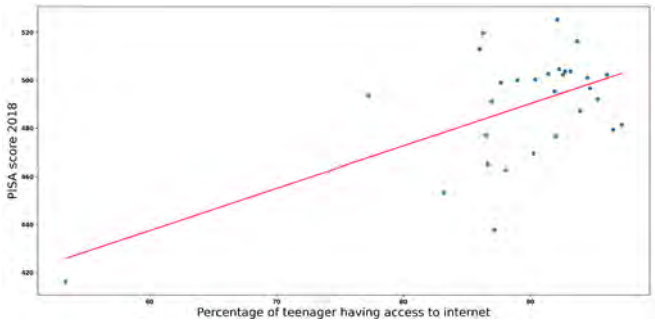


Figure 1. *PISA score 2018 against the percentage of teenagers with access to the internet*

A clear positive correlation is obtained by plotting the PISA score in 2018 of 30 countries against the percentage of an average teenager having internet access that year. The Pearson Correlation Coefficient of y against x is 0.59020708, indicating a moderately strong correlation between teenagers accessing cyberspace and their PISA score.

3.4 The Correlation Between the Internet Freedom Index and the Human Development Index

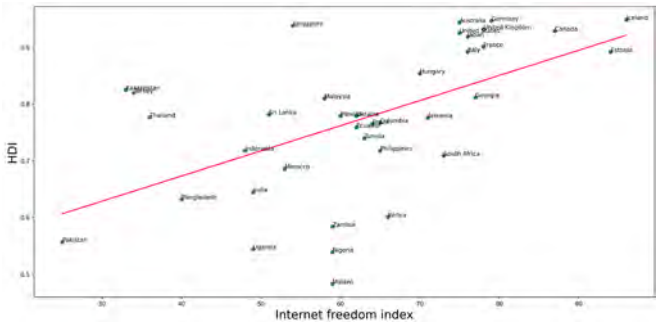


Figure 2. *Human Development index against internet freedom index*

This best-fit line shows that countries with less cyberspace freedom enjoy a lower Human Development Index than countries with more cyberspace freedom. The Pearson Correlation Coefficient of y against x is 0.55478705, which indicates a

moderately strong correlation between cyberspace freedom/cyberfreedom and a country's Human Development Index.

3.5 Internet freedom against the Gross Domestic Product

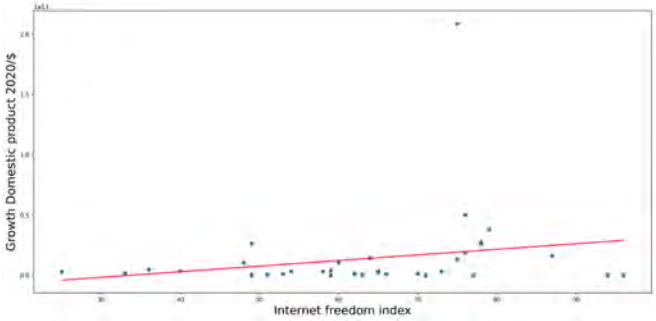


Figure 3. *Gross Domestic Product against internet freedom index*

This best-fit line shows that countries with less cyberspace freedom enjoy a lower Gross Domestic Product than countries with more cyberspace freedom. The Pearson Correlation Coefficient of y against x is 0.21051549, indicating a weak positive correlation between the internet freedom index and the Gross Domestic Product. This result reflects that cyberspace freedom drives only a small factor of a country's gross domestic product. Countries like Russia, with many natural resources, often offset the censorship's effect on human development and the state's economic development.

3.6 Conclusion of correlations

Since both Gross Domestic Product and Human Development Index's correlation with the internet freedom index is positive, a state with more cyberspace freedom will have more potential for development and a faster development speed. Also, teenagers with more internet access have stronger academic abilities.

This correlation begs the question, why, if it is the case that more cyberspace freedom correlates to a higher human development index, better economic prospects, and superior academic abilities, do all countries not follow this model?

4. Excuses Countries Have About Restricted Cyber-freedom

4.1 Cybercrime

There are many excuses governments use to cover up their restrictions on cyber freedom, be it preventing election day chaos, helping restrict access to pornography, protecting citizens from capitalism, and reducing teenage gaming addictions. In this research paper, I will analyze the main excuse—cybercrime.

Like any realm, virtual or real, that man dwells in; crime is a problem in cyberspace. However, in the cyber realm, crimes are hard to track, easier to perpetrate, and difficult to spot. Thus, countries often excuse their monitoring of cyberspace and restrictions on their citizens' cyber freedom to protect cyberspace

from crimes. Accordingly, this portion of the research paper will address the claim: does more cyber freedom correlate positively with more cybercrime?

One of the most prominent types of cybercrime is phishing attacks. According to NCSC, Phishing is when attackers attempt to trick users into doing 'the wrong thing,' such as clicking a bad link that will download malware or direct them to a dodgy website *(Phishing Attacks)*.

To compare the number of cybercrimes in a country with cyberspace freedom, I use the data from Kaspersky Lab users whose devices triggered Anti-Phishing out of all Kaspersky users in the country in 2021 *("Phishing")*.

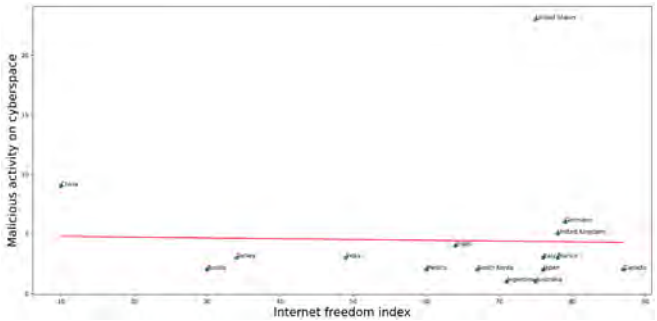


Figure 4. *Malicious activity in a country's cyberspace against the internet freedom index*

This best-fit line shows that countries with less cyberspace freedom have a higher share of malicious computer activity.

The Pearson Correlation Coefficient of y against x is -0.02631801, indicating a weak negative correlation between the internet freedom index and cybercrime. This result may be a surprise, but countries have always known that restricting people's cyberspace usage does not correlate with less crime in cyberspace. This statistic is understandable in real life as cybercrime can be conducted in intranets and splinternets, similar to cybercrime in global cyberspace. Countries who allege their censorship and control to be legitimate by indicating they restrict cybercrime are statistically incorrect and have no ground to stand on.

4.2 Regime Stability Against Cyberspace Freedom

I define political regime change as changing the party with majority control over a government (some European nations do not give all the political power to one party and divide seats between people).

4.3 Data Collection

- Estonian election results *(Nohlen and Stöver)*
- Egypt election results *(Egypt | Jadaliyya)*
- *(Psephos - Adam Carr's Election Archive)* for the United States, Canada, China, Germany, Japan
- Uganda election results *("HISTORY OF ELECTIONS IN UGANDA")*

I obtained a political regime change database by collecting all the above data and manually inputting the number of political party changes and the period of a country's existence. [1]

4.4 Data Correlation

I generated a database from the above data, exemplified in the figure below. The frequencies of change are calculated by the number of changes in a political party divided by the total years of the country's election history.

Table 1. *Data on political party change in countries*

	Country name	frequency of chang	change in political party	year	start year	end year	total years of election
0	United States	0.115854	19	1857-2021	1857	2021	164
1	Canada	0.116883	18	1867-2021	1867	2021	154
2	China	0.009174	1	1912-2021	1912	2021	109
3	Germany	0.140625	9	1949-2013	1949	2013	64
4	Japan	0.034483	2	1955-2013	1955	2013	58

A scattered graph with the best-fit line could be plotted by the frequency of political change against the internet freedom index.

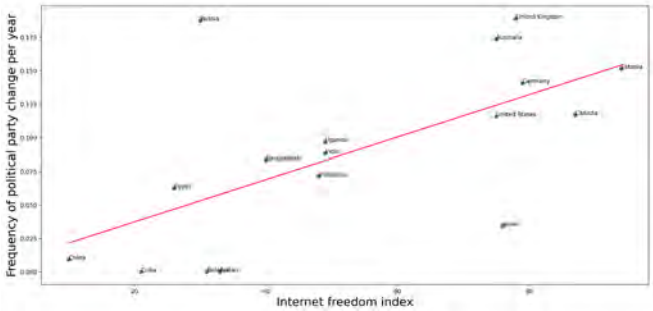


Figure 5. *Frequency of change of political party against internet freedom index*

The Pearson Correlation Coefficient of y against x is 0.62553849, indicating a moderately strong positive correlation between the frequency of political party change per year against the internet freedom index.

4.5 Irrationality of Governments

The above section shows that a higher internet freedom index correlates with higher regime change frequency. However, these changes are not violent takeovers and are more likely smooth transitions. For example, apart from rare cases of individual selfishness in the United States, a Democrat government changing into a Republican government or vice versa is expected, without ever making the nation descend into revolutions or uprisings.

1. See table in appendix

However, exceptional cases could be made for countries like Iran and Haiti, who argue that freer cyberspace may cause more political instability, further diminishing the country's development potential.

4.6 Examining Overall Political Stability in a Country and Its Relationship with Cyberspace Freedom

Although countries are trying to use cyberspace censorship and other restrictions to improve their hold on power, there needs to be more examination of how cyberspace freedom influences the political stability of a country. Here, political stability is positively correlated with citizens' satisfaction and inversely correlated with the number of political protests, with other relevant factors also considered.

Data gathering: to examine the relationship between the political stability of states and their cyberspace freedom, data about political stability is gathered from the database *"Political Stability by Country, around the World."* After selecting data and ensuring all countries with political stability data also have internet freedom index values, data from thirty-six countries are plotted against their internet freedom index.

The graph consists of scatter points of all the countries; the best-fit line is obtained and plotted in the below figure in red.

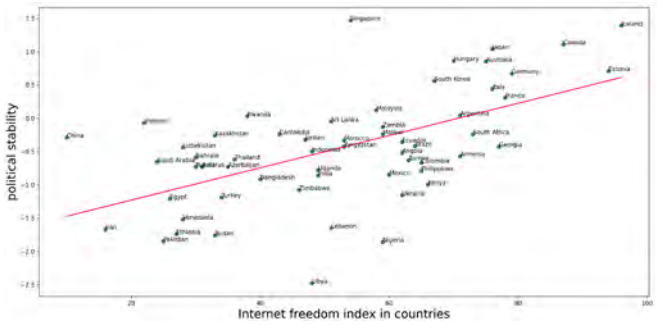


Figure 6. *political stability against the internet freedom index*

The Pearson Correlation Coefficient of y against x is 0.63012748, indicating a moderately strong positive correlation between political stability per year against the internet freedom index.

The correlation coefficient and the best-fit line show that countries with greater cyberspace freedom enjoy a more politically stable country. States and political parties should not fear higher cyberspace freedom causing them to lose power, because more cyberspace freedom helps the country's political stability increase.

5. Case Studies

5.1 Measuring Online Sentiments Toward Countries

Reddit is a social media platform with various discussion boards allowing people to discuss relevant information on particular topics. Apart from being an outstanding example of cyberspace freedom, it is an excellent tool for measuring internet

opinions. People on this site are usually frank (occasionally overly frank) with their opinions due to their ability to stay anonymous.

The subreddit I am examining is r/worldnews, a discussion board filled with opinions about countries and world leaders.

With the PSAW python module and a Reddit API, I crawled and analyzed 4,327,411 entries and labeled them with the subject and the commentator's attitude towards it. Using Natural Language ToolKit and a python script, I turned my Raspberry Pi into a data receiving center. I kept it running for more than 294 hours, collecting data from 2012-2022 from about 102 countries. The data is invaluable for analyzing censorship's effect on countries' online reputations

5.2 The dataset from Reddit

After collecting data from the Reddit website, I parsed it into a "pandas" DataFrame. The Python "pandas" module is a helpful plugin with data manipulation, normalization, and exporting features. The following figure shows the shape of the database.

Table 2. *Reddit dataset before labeling*

	title
0	title
1	EU to invest 300 billion euros in green energy...
2	Trump Bragged He Had 'Intelligence' on Macron'...
3	1M Russians Enter EU Since Ukraine War Start, ...
4	Taiwan's working-age population projected to h...
...	...
157000	[deleted by user]
157001	Who is the good baby at home?
157002	South African doctor who first alerted authori...
157003	[deleted by user]
157004	Georgia may get looser Covid rules because loc...

157005 rows × 1 columns

5.3 Data Sorting of the Reddit Data

Since many Redditors do not necessarily follow the rules of the subreddit, compiling a database without filtering through its content is costly to the accuracy of the conclusions. Thus, I processed all the sentences only to leave ones with a politician or a country name as its subject. After sorting through the subreddit data, each post's subject is identified.

Table 3. *Reddit dataset after labeling*

	title	neg	neu	pos	compound	label	subj
0	title	0.000	1.000	0.000	0.0000	neu	None
1	Mark Ruffalo Confirms What We Suspected All Al...	0.137	0.863	0.000	-0.2263	neg	Ruffalo
2	Jharkhand: Ankita Kumari, a 12-std. student ha...	0.255	0.703	0.042	-0.9325	neg	Shahrukh
3	Mayor urges residents to flee ahead of rising ...	0.000	1.000	0.000	0.0000	neu	Mayor
4	'Get out now': Mayor urges residents to flee a...	0.000	1.000	0.000	0.0000	neu	Mayor
...	...	...	...	...	...	...	...
37949	The 12 times Texas police have changed their s...	0.177	0.823	0.000	-0.6486	neg	police
37950	Trump: US should fund safe schools before Ukra...	0.000	0.756	0.244	0.4404	pos	US
37951	New independent Mexican union wins wage increa...	0.000	0.571	0.429	0.7184	pos	union
37952	Kurt Cobain's 'Smells Like Teen Spirit' Guitar...	0.000	0.682	0.318	0.4939	pos	None
37953	Zelenskyy to Attend G20 Summit Virtually if Ru...	0.281	0.719	0.000	-0.5994	neg	Zelenskyy

37954 rows × 7 columns

5.4 Dataset Overlook After Language Analysis

After obtaining the title's subject, an array of countries and attitudes are made to collect the total attitude of online people towards the country. However, as people often address their opinions towards particular politicians in a country that controls the state, it is vital for the accuracy of the data that a post mentioning a country's ruler/leader also contributes to the attitude_coefficient of a country.

To attribute all online attitudes of politicians to their respective countries, I used the helpful database Political Leaders' Affiliation Database *(PLAD)* from Harvard's database collection *(Dreher).*

Table 4. *Database overlook of PLAD*

	idacr	leader	plad_id	archigos_id	startdate	enddate	startyear	endyear	adm0	adm1
0	AFG	Najibullah	AFG_1986_1	8243611b-1e42-11e4-b4cd-db5882bf8def	04may1986	16apr1992	1986	1992	Afghanistan	Paktya
1	AFG	Mojadidi	AFG_1992_1	8243611c-1e42-11e4-b4cd-db5882bf8def	28apr1992	28jun1992	1992	1992	Afghanistan	Kabul
2	AFG	Burhanuddin Rabbani	AFG_1992_2	8243611d-1e42-11e4-b4cd-db5882bf8def	28jun1992	27sep1996	1992	1996	Afghanistan	Badakhshan
3	AFG	Mullah Omar	AFG_1996_1	8243611e-1e42-11e4-b4cd-db5882bf8def	27sep1996	13nov2001	1996	2001	Afghanistan	kandahar
4	AFG	Hamid Karzai	AFG_2001_1	8243611f-1e42-11e4-b4cd-db5882bf8def	22dec2001	29sep2014	2001	2014	Afghanistan	Kandahar

5 rows × 41 columns

For this research paper, I will only use the "leader" and "adm0" columns, which include data for politician names and country names.

With this database, any comments about world politicians contribute to the politician's country's online reputation. For example, if one comment negatively on the subreddit about "Putin," it will be counted as a negative attitude towards the state of Russia.

5.5 Sentiment analysis of Subreddit Submitted Titles

NLTK utilizes machine learning to calculate the approximate attitude value of a sentence, with "1" being entirely positive and "-1" being completely negative. However, the model's return value is always a fraction between the two extremes. Thus, I divided the attitude values into three categories for easy calculation.

Sentences with attitude values over "0.2" are considered positive comments.

Sentences with attitude values lower than "-0.2" are considered negative comments.

Sentences with attitude values between "0.2" and "-0.2" will be considered neutral comments.

I have compiled a database of online attitudes towards countries from 2022/2/3-2022/8/3. The database compiled is in the appendix.[2]

5.6 The Case of Russia

Russia is the stereotypical surveillance state, with every move of its citizens controlled and put under heavy scrutiny. One wrong move in cyberspace and one could be taking a permanent vacation with a stab wound or a bullet hole in one's heart, and based on the severity of one's "crime" against the government, one may see one's family joining them (Mayer).

Recently, as the Russian invasion of Ukraine unfolds and Russia consistently embarrasses itself with its subpar military, the Russian government is implementing a new heavy-handed internet censorship law.

Russia clamped down harder on news and free speech in March of 2022 than at any other time in Vladimir Putin's regime. Putin's new law includes blocking access to Facebook and major foreign news outlets and enacting a law to punish anyone spreading "false information" about its invasion of Ukraine with up to 15 years in prison *(Troianovski and Safronova)*. For example, Alexei Gorinov, 60, was arrested in April after he was filmed criticizing the invasion in a city council meeting *("Russia-Ukraine War")*.

From my Reddit analysis data, each negative value is considered as a "-1" value to the country's online reputation, a neutral comment is a "0" value, and a positive comment is a "1" value. Summing the total reputation score of Russia shows the country's online reputation in that particular month.

Table 5. *Online attitude towards the state of Russia*

Date	2022/2—2022/3	2022/3—2022/4
Russia	-0.00194	-0.06622

I calculate a country's reputation by the cumulative internet reputation value divided by the number of comments people make in the subreddit. Although Russia attempted to better its online reputation through harsh punishments, cyberspace is often too elusive. The attempt to restrict cyberspace freedom resulted

[2] Check appendix for full table

in a dramatic spike in negative comments about Russia. Considering the seven hundred thousand plus posts made during that month, the change from attitude_coefficient of -0.00194 to -0.06622 reflects an increase in tens of thousands of negative commentaries, reflecting a significant increase in unfavorable opinions towards Russia. Russia has a cyberspace population of 146,069,910 *(Russia Population (2022) - Worldometer),* and assuming r/worldnews represents the Russian people's attitude towards Russia, during the one month between February to March, Russia gained an additional 9,389,374 negative comments. The data shows that this Russian attempt to restrict cyberspace dissent through censorship has backfired.

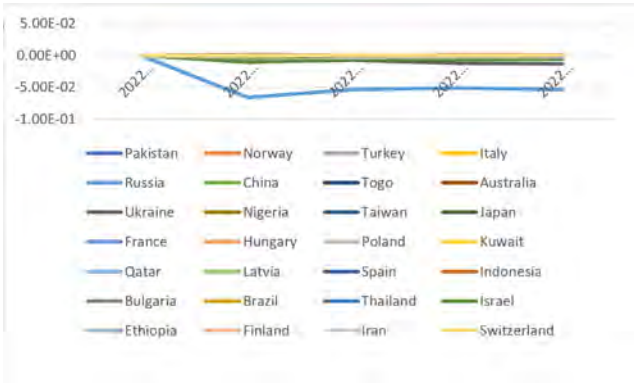


Figure 7. *States' cyberspace reputation over time*

In the above figure, states' cyberspace reputation over time is shown. The blue line below all other countries shows Russia's cyber reputation dramatically plummeting after it invaded Ukraine. Although the Russian government implemented various cyberspace freedom restriction laws, the downward trend of Russian cyber-attitude has continued and has even increased. In this case, the Russian cyberspace freedom restriction law is not successful in stopping online dissent.

5.7 The Case of China

Three days after a vicious attack on a group of women in China, the Chinese state-controlled media Weibo announced a zero-tolerance policy toward users who spread "harmful speech," including comments that "attacked state policy and the political system" or that "incited gender conflict" *(Zhang).* The decision, however, also influenced China's reputation in cyberspace. The figure below illustrates China's online reputation change from before the incident to after.

Table 6. *Countries and their online reputation - country: China*

Time	2022/06-2022/07	2022/07-2022/08
attitude_coefficient	-0.003113501	-0.004072656

Even though the new cyberspace restriction guideline on Weibo should stifle cyberspace discussion about this particular Chinese incident, there is a 0.000959155 increase in negative cyberspace comments about the People's

Republic of China. China currently has a cyberspace population of 1,456,487,861 *(China Population (2022) Live — Countrymeters)*. Assuming that people's attitude towards China on the subreddit reflects Chinese citizens' attitude toward China, there is a corresponding surge in 1,396,998 more negative comments about the Chinese government within China.

Punishments for online behavior may deter individuals from engaging in inappropriate actions. Instead, they can suppress free speech and drive individuals to use alternative platforms that the government may not regulate. This can harm a country's reputation in the online world and potentially lead to the spread of false or harmful information. Instead of relying solely on punishment, educating individuals about responsible online behavior may be more effective and provide people with tools to safely and effectively express their thoughts and opinions.

5.8 Thailand - A Pleasant Exception

In numerous figures presented in this research paper, Thailand always seems to be the exception. Thailand scores low on the internet freedom index yet consistently ranks among the top of the Human Development Index, political stability, and domestic product growth. As a country ruled by a parliamentary constitutional monarchy on the surface and a military government, Thailand somehow still scores high on many indicators of citizen happiness and state development. I will explain this discrepancy through the following reasoning.

Firstly, Thailand is a constitutional monarchy with the power to delegate power but not to exercise powers directly. Monarchs in Thailand reign but do not rule. A central figurehead is one of the reasons Thailand has "succeeded" in censoring its people without having massive uprises and invoking the wrath of its people, like in states like Russia and China. One reason could be that Thailand masks political suppression as protecting the king's name. Citizens even actively report each other when others violate the king's authority or speak ill of the political system. The Seri Thai Movement, a pro-monarchist political group, encourages its 80,000 members to defend the monarchy by reporting lese-majeste violators through its Facebook page and the popular Seri Thai Web Board *(Sinpeng)*. The devotion-to-king sentiment legitimizes Thailand's effort to repress dissent and political disagreement. Moreover, Thailand seldom puts travel restrictions on even its worst critics, which means people who disagree with the government could quickly leave the country and seek a more democratic and free place to live. The ease of leaving helped Thailand ensure that most of its population was made of royalists and people who agreed with the regime.

Secondly, Thailand has a partially-free cyberspace. As a visitor, one would have access to almost all of their normally visited websites and social media. Websites like Reddit, Facebook, and Twitter are all allowed in the country, except those parts that oppose the monarchy. By threatening to remove companies from its territory entirely, Thailand successfully negotiated with sites like YouTube, Facebook, and Twitter to make anti-government pages inaccessible to people with Thai accounts. This partial freedom stops Thailand's citizens from striving to have the same access as people outside the country. For a citizen in China, Facebook would not be something they could access without a private network, which creates plenty of mystique around these "foreign websites." Humans explore mystery-surrounded things, and research indicates that reducing uncertainty about a subject is a significant motivator for human exploration *(Ten et al.)*. Completely restricting access to a particular social media will cause curiosity and mystery. In contrast, a

partial block of a social media site leaves citizens none the wiser, and fewer Thailand citizens would revolt because their favorite streaming site is blocked.

Last but not least, Thailand exerted its power of normalization in its cyberspace. By policing what is "right" and "wrong" behaviors in cyberspace, the Thai government normalizes cyber behaviors that it approves. In 2010, the Cyber Scout program in Thailand recruited 100,000 scouts to become "cyberwarriors" and defend the Thai king's reputation in cyberspace. Cyberwarriors report and punish "bad" behaviors in cyberspace while praising and quoting "good behaviors" on government websites. Thailand wishes to normalize the population in cyberspace to a royal, approving mass that worships the king. Although this type of cyberspace control still violates the fundamental freedom of Thai citizens, there is no denying that this type of censorship works for Thailand.

5.9 How Freedom Pays—New Zealand

New Zealand is a historically free country in terms of cyberspace and other domains. However, New Zealand is still not a free-for-all cyberspace domain. Censorship in New Zealand is governed by the Films, Videos, and Publications Classification Act 1993 and associated regulations *(MSD)*. Unlike other countries, New Zealand's cyberspace freedom restrictions are all about helping fight criminals who either violate public decency, personal privacy, or other people's intellectual property. New Zealand labeled all of these inappropriate cyberspace materials as objectionable publications.

New Zealand's internet is open to any political discourse as long as it does not violate other people's rights. This freedom in cyberspace allowed New Zealand to be ranked at the top of the world's cyberspace freedom index list.

Below is the internet attitude_coefficient for New Zealand from 2022/03/02-2022/08/02

Table 7. *New Zealand attitude_coefficient database 2022*

New Zealand	1.32E-05	8.14E-05	4.04E-05	2.63E-05	1.86E-05

New Zealand is one of the only countries that have a positive attitude_coefficient in cyberspace where negative comments are prevalent. By giving New Zealanders freedom to discuss politics in cyberspace, the New Zealand government not only earned positive comments in hostile and antagonistic cyberspace but also ensured that New Zealand will be remembered as one of the countries with the highest freedom values in cyberspace.

6. The Ultimate Game

While countries and municipal bodies consider themselves unique due to their economic prowess, rich cultural background, or remarkable technological advances, they are individual agents in a game–the game of world politics.

The game of politics, in its essence, is that of strategies and wits. By changing various governing policies, states can navigate the game of politics and eventually come up on top. The real world of politics is also interlaced with war, violence, and persecution, making the consequences of wrong policy decisions extremely risky.

In this research paper, I will simulate the game of politics using deep-learning neural networks. Neural networks are made up of "neurons" in layers that can connect and be assigned a value to determine how the output from a neuron is respective to the input to the neuron. Thus, a decision-making network and artificial intelligent country agent could be simulated by constructing a net of neurons.

With the programming language of Python and packages Matplotlib, Keras, NumPy, math, and random, I can create a program that generates N numbers of country agents and assigns them each to a simple neural network. Then, set rules for war, growth, political rebellions, and dissolution of large nations.

6.1 Neural Network Architecture

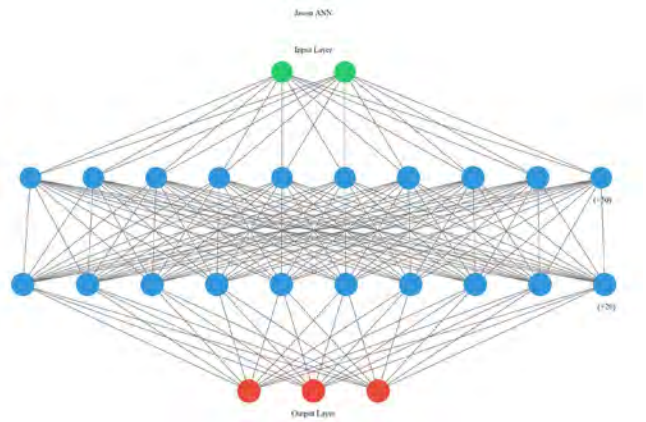


Figure 8. *DNN network visualized with graphiz*

To simulate a country's decision-making ability, I used a four-layered Deep Neural Network (DNN) with two inputs, 60 neurons in the first hidden layer and 30 neurons in the second hidden layer, and three output neurons in the output layer. The input will be the location of all other countries and their respective freedom index. The output will be the country's acceleration on the x-axis and the y-axis and the freedom setting of the country. For this research, the assignment of 90 neurons in hidden layers may seem inadequate, but noting the infinite number of possibilities for assigned value for each neuron and the complexity of the neuron connections, the neurons are enough for simulating complex decisions made by states.

I selected FFNN *(Feedforward neural network)* as my neural network architecture. FFNN is a simple network with input, hidden, and output layers. The combination of neurons in each layer will propagate forward to the next layer. It then goes to the activation function, and outputted values and their combinations are the inputted values of the next layer. It will continue for all layers, up to the last layer that gives the final outputted values *(Ghayoumi)*.

I utilized an unsupervised learning model called the genetic model in this paper. Since it is impossible to give feedback to the deep learning model without having a subjective judgment on what is necessary for the countries' development, an unsupervised learning model needs to be used. The genetic model automatically deletes any country with a fitness value too small or after being destroyed by another country in war. After the country's population becomes too low, new countries will

be generated with the genetic construction of countries with the highest fitness value. Thus, only the countries that can maintain high fitness and sphere of influence will remain in the simulation after every generation.

6.2 Parameters for the Countries

- x, y—the x and y coordinate of the country agent; these values do not have an initial value
- free—the freedom index of a country; this value is initialized as 0 at the start of the program
- turmoil—the amount of political turmoil in a country; this value is 0.1 at the program's start. This value is calculated inversely proportional to the internet freedom index
- v, dv—the velocity of the country and the acceleration of the country
- fitness, size—the fitness and size of the country agent; these values are initialized and adjusted according to the program. Any agent smaller than 0.01 will be too insignificant to continue its country's development and will be removed from the countries list—simulating the destruction of a country in real life.

6.3 Movements of Countries

While it is possible to simulate countries having conflict by providing the means with which to attack each other, such a simulation would be resource exhaustive, and the results would not be easy to sort through. Thus, I am representing the country's decision to be peaceful or violent by allowing them to move in an arena-like setting. When two spheres–representing the sphere of influence of the countries–conflict, the one with less fitness will be killed (removed from the map); however, the larger country will suffer damage in both its fitness and size. Since countries control their movements, they can choose whether to be peaceful and develop or be violent and eliminate enemies.

6.4 Rules Inherent to the Game of Politics

- Rule of birth: settings['pop_num'] countries will be generated randomly with a random neural network.
- Rule of death: any country that moves out of the observable arena will be eliminated from the entire country array. Any country with negative or 0 fitness will also be killed/eliminated.
- Rule of war: every n second, when countries are close enough, they will fight, and a country with higher fitness value will survive.
- Rule of space occupation: no two countries can occupy the same position without being eliminated.
- Rule of turmoil: a country with a turmoil index higher than three will suffer a revolution and be divided into two smaller countries. This particular rule reflects the effect of too much political instability in a country, leading to the country splitting, like the "velvet revolution" in which Czechoslovak split into Slovakia and the Czech Republic *(Rychlík)*.

6.5 Peaceful Model

In a perfect world, without any external conflict, countries could develop whatever strategy they deem beneficial to their country. If a country decides to have absolutely no internet freedom/freedom overall, they can do their slow controlled growth; on

the other hand, if a country embraces freedom and allows its citizens to do whatever they want without any restrictions, there will be risks that the country will collapse due to brain-drain and political dissent. Still, there is also a chance that the country will become extremely prosperous and develop into a major world superpower.

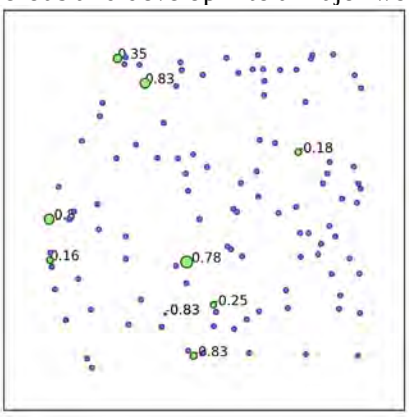


Figure 9. *An initial state for the country simulation*

A frame close to this will be displayed at the start of the simulation. The digits labeled in the above figure show each nation's freedom indexes. These are based on decisions individual agents have made.

In these previous simulations, a country will be randomly generated at a location in the simulation. The only variable the individual neural networks could adjust is the state's cyberspace freedom.

6.6 Visualizing the freedom index of each nation

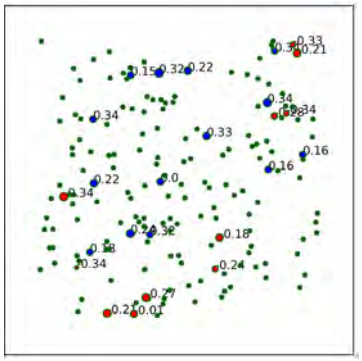


Figure 10. *A visualization of the freedom index in countries*

Plotting countries now on an x and y-axis will show how free a country is. A country with a freedom index smaller than 0.2 will be plotted as red, which indicates that it is pretty authoritarian; on the other hand, if a country has a freedom index higher than 0.2, then it will be plotted as blue, which indicates it is quite democratic and has a high internet freedom index.

6.7 When no penalty is given for providing less internet freedom

In a system without consequences for countries having less internet freedom, countries will adapt to having less internet freedom. To test the neural network's adaptability, I will run the following simulation mutation through my Linux server: no penalty for less internet freedom, political turmoil is disabled, and countries cannot move

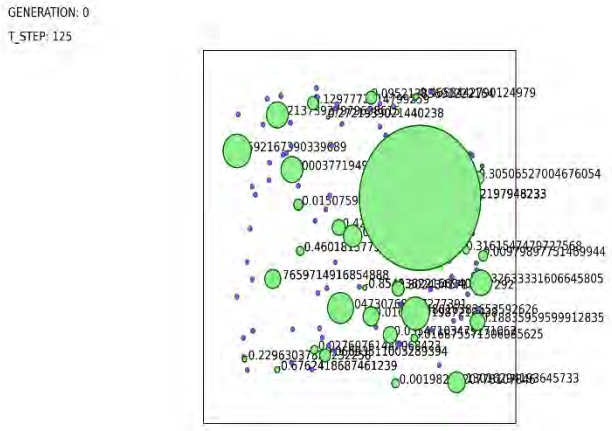


Figure 11. *When no penalty is given for providing less internet freedom*

Unsurprisingly, countries naturally choose to be unfree when no reward is given for having a higher cyberspace freedom value. The dominant country has a freedom index of -0.0748567. The model has proven to be adaptive and able to change rapidly according to the scenario it is subjected to.

6.8 Short-term Strategy—Peaceful

To simulate short-term strategy for countries, I temporarily ignore the influence of a country's freedom on its utilization of resources. While more political turmoil will result in a country being more likely to dissolve, it does not influence the country's development or ability to utilize resources effectively.

The average freedom for the succeeded country is -0.070771 (succeed is defined as size>2.0). Once the country's size exceeds 2, they are considered a superpower and will take over the political ecosystem. Countries choosing a no-freedom strategy gain a significant advantage in the short run as it reduces brain drain from their country and helps them quickly develop their country's sphere of influence using the resources they obtain.

Here are some case studies from the simulations where situations leading to a country becoming a superpower (defined as size>2.0, fitness>25) are analyzed.

6.9 Case Study #1

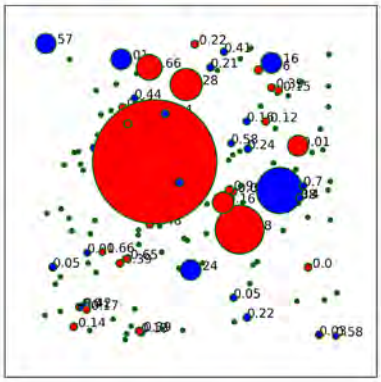


Figure 12. *Internet freedom index -0.8529945778830136*

This country adopted a no-internet-freedom policy. Due to the lack of competition around and the low brain-drain rate, it quickly dominated the map of countries, engulfing nations in the process.

6.10 Case Study #2

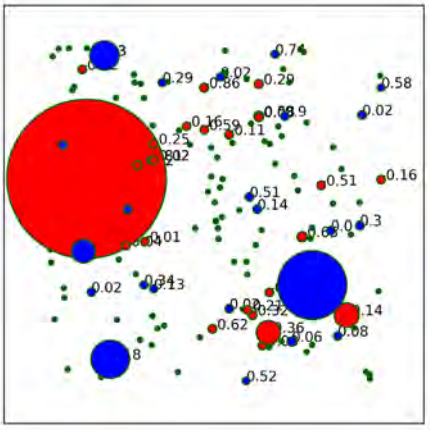


Figure 13. *Internet freedom index -0.5785048801657333*

This country adopted a no-internet-freedom policy. Due to the lack of competition around and the low brain-drain rate, it quickly dominated the map of countries, engulfing nations in the process. However, after a few pieces of training, other countries learned to develop quickly using more freedom.

6.11 Case Study #3

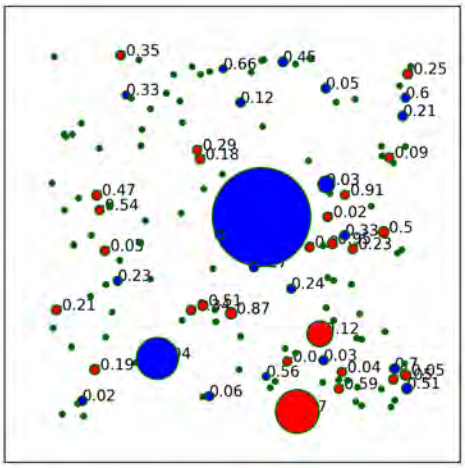


Figure 14. *Internet freedom index 0.27*

An example of a democratic freedom country rising to a superpower; it seems that choosing more freedom is also a viable way to develop in peaceful times.

6.12 Long-term Strategies—Peaceful

In the long run, more freedom nurtures creativity, which makes resource use more efficient, causing faster development. The model ran with a shorter t_step (time step) reflects this efficiency effect, and freedom's effect on the resources is re-established.

This time, the table is completely flipped, as neural networks deemed more internet freedom as the safer and faster way to develop in the long run.

Table 8. *All countries that achieved world superpower and their internet freedom index*

.950607	.973711	.695152	.484669	.34659	.865344	.995646	.813098	.95686	.359713
.980455	.861199	.300697	.741641	.973004	.973616	.849363	.998577	0.95904	.552106
.986846	.894574	.513648	.938662	.652715	.915882	.773806	.929558	0.979713	.979109
.870552	.99969	.701133	.834845	.697355	.860565	.67431	.791292	0.774526	.962477
.805753	.991573	.828273	.848947	.696014	.535602	.293788	.895692	0.92064	.975992
.963079	.676252	.634341	.595066	.998508	.985157	.943817	.90096	0.805214	.28694
.989952	.996815	.42987	.7654	.857763	.875123	.769319	.981131	0.933431	.310707
.760472	.989653	.887461	.271826	.88568	.986298	.883826	.929441	0.380752	.45113
.510951	.53617	.597076	.677389	.906999	.596275	.342715	.935311	0.835239	.71073
.337333	.679815	.852465	.561231	.842476	.996109	.995482	.481207	.966347	.751431

The average internet freedom index is 0.77392796, which reflects a moderate-to-high freedom index for countries that successfully win the game of world politics.

6.13 Long-term Strategy—with War

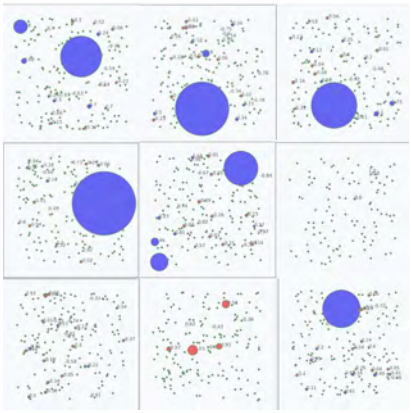


Figure 15. *Different scenarios of this setting*

In this simulation, to simulate states' decisions to be aggressive toward other states, the individual agents are allowed to accelerate in the x, y direction. Once two countries touch, there will be a simulated war where the two states compare their fitness values, and the fitter country will survive the war. States use up their fitness value when they accelerate in any direction, and war reduces the state's fitness value by ¾ of the other state's fitness value.

The average internet freedom value for countries succeeding in becoming world superpowers is about 0.5398. Compared to the model with peace, countries are more cautious in changing their freedom to too high, as political turmoil in the country coinciding with a war with another country will cause the state to be destroyed.

Another fascinating trend in the long-term simulation of war is that countries are more polarized than average in a short-term situation, meaning that countries tend to have either extremely high or extremely low internet freedom.

In all 100 simulations, there are 14 scenarios where countries with opposite cyberspace freedom indexes clashed in war and were mutually annihilated. These scenarios reflect the real dooming possibility of a global war sparked by differences in online freedom. If freedom allows a country's citizens to escape to a neighboring free country, the first country's economy will be devastated, which may lead to war between the states to prevent further brain-drain.

Table 9. *Three outputs from the program under 5.6 situations (the values are for countries' freedom values)*

> GEN: 0 BEST: 0.9999925840971187 AVG: 0.0114602712194514 WORST: -0.9978294144112
> GEN: 1 BEST: 1.039904097940312 AVG: -0.17319531626204002 WORST: -0.9534011505134
> GEN: 0 BEST: 0.9999874176843824 AVG: 0.09697703903730813 WORST: -0.9998677868578

6.14 Long-term conflict model with states unionization

A new rule is made for the new simulation model, where countries with similar freedom indexes can unionize and form a more significant state with combined genetic information for the neural network. This new rule gives countries incentives to consider their freedom value similar to real life, as it is easier to form alliances with countries if they have similar policies.

After simulating 1000 generations, the countries' behaviors stabilize, and a pattern for state behaviors is formed. When two neighboring states notice they have opposite freedom values, more than 70 percent of the time, the two countries immediately go to war. When two states are close and have similar freedom values, they tend to merge and form a stronger country with higher development speed.

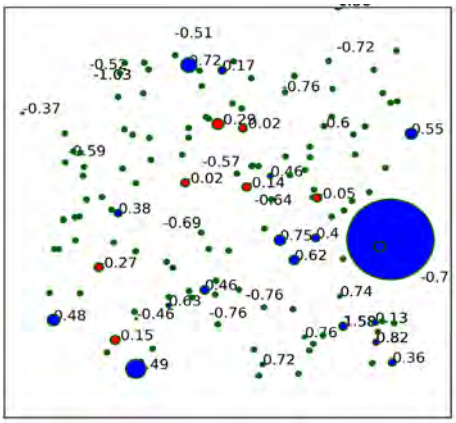


Figure 16. *Countries in unionizing model*

The average cyberspace freedom index value for states is 0.7693309. Thus, despite the new unionization rule, the neural network still deems allowing more freedom in cyberspace as the optimal strategy.

When unionizing to become a larger country is an option for states, countries have higher freedom index value than when countries do not have unionization as an option. In this simulation mode, countries are more likely to become freer since it is easier for them to form a larger union to thwart any war attempt by other nations that can have rapid development in a short time. Also, I observed that when states with higher cyberspace freedom surround states, the surrounded states increase their freedom values. When states with lower cyberspace freedom surround states, the surrounded states decrease their freedom values.

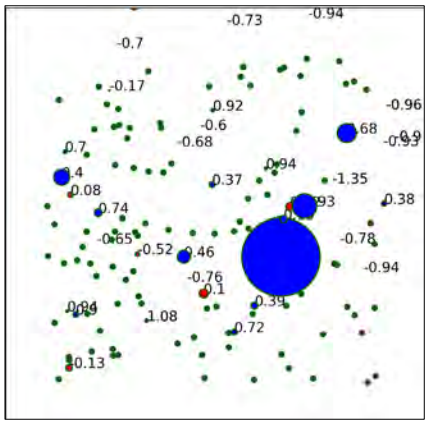


Figure 17. *Countries in unionizing model*

In the middle of this simulation, states all have less than 0 cyberspace freedom values. However, once the two large blue states start expanding quickly, the smaller states start increasing their cyberspace values to be seen as "on the same side" by the stronger state and hopefully to be able to merge with the stronger state and contribute its own identity to a larger union (in terms of the model, adding its genetic information to the larger union.)

This new simulation revealed another essential factor in a country's cyberspace freedom – its surrounding political environment. If countries with loose cyberspace control surround a state, restricting cyberspace usage can cause tension between states and their citizens to seek more freedom. On the other hand, when surrounded by countries with iron grips on their citizens' cyberspace use, a state with high freedom values in cyberspace can antagonize the neighboring states.

Also, in this more realistic simulation, the polarizing power of cyberspace freedom is shown. After 50 generations of simulations, a clear pattern of ideology division is shown. Countries form their region of similar cyberspace freedom values, and countries with opposing cyberspace freedom values are quickly annihilated or assimilated. After several hundred time-steps, countries left are all grouped in different unions of similar cyberspace freedom values, and movements of the states slow to a stop.

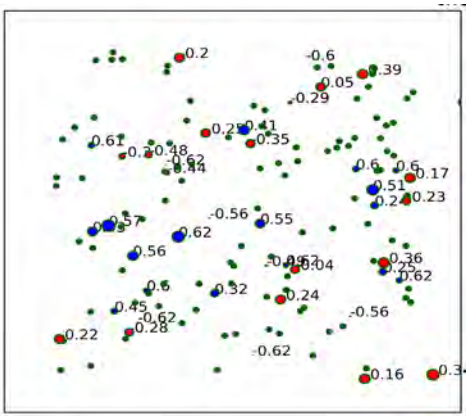


Figure 18. *Countries in unionizing model*

Another observation about countries in the unionizing model is that the movements of countries are slower than in other models. Since movement in these simulations mirrors conflicts in real life, the reduced movements show the reduced aggression between countries when unionizing is an option. Countries are relatively safer when they can form a larger union instead of competing with every other country.

Out of the 100 simulations, only 12 scenarios ended in the total annihilation of all the states. In 4 other scenarios, one union of nations successfully overpowered the other union of opposite cyberspace freedom values, three of which were nations with cyberspace freedom values over 0.5. The other simulation showed a union of nations with a cyberspace freedom value of -0.9834 taking over the arena.

7. Discussion

Cyberspace is a shared space for people to gain knowledge, express themselves, and connect. However, some governments are unwilling to relinquish control of this new frontier. Although some successful cases of cyberspace freedom restrictions may exist, statistics reflect that, for most states, cyberspace restrictions are harmful to the stability of the nation and the future development of the state.

The first part of the research illustrates a positive correlation between teenagers' access to cyberspace and academic performance in internationally standardized tests, demonstrating the positive effect of high cyberspace freedom on an individual level.

Correlation between the internet freedom index and Human Development Index, Gross Domestic Product, Teenager Academic Performance, and Political Stability also shows a positive correlation between cyberspace freedom and a country's development potential. The only inversely proportional value to the internet freedom index is regime stability, which explains states' reluctance to let their citizens have free cyberspace activities. However, the later part of the research illustrates how countries should not only look at regime change frequency because overbearing in cyberspace could have worse consequences than a regime change.

Through online attitude analysis, I found that countries increasing censorship during a particular period will cause negative online attitudes towards that country to surge. Thus, countries should not utilize any sudden increase in censorship to reduce online dissent.

Last but not least, through neural network simulations, I concluded that countries need to relinquish control of their citizen's cyberspace usage to win the ultimate game of world politics. A country's surrounding nations' cyberspace policies also influence the country's cyberspace freedom value choice. Cyberspace freedom restrictions proved to be a potent cause for global conflicts, and states should heed their actions in controlling cyberspace as real-world consequences can be imminent and devastating.

It is alarming to consider that world leaders are aware of the potential dangers and issues brought by restricting cyberspace freedom, yet continue to ignore the adverse effects of maintaining their hold of power on the country through cyberspace. From more political turmoil to even all-out war between nations, the consequences of such actions can be devastating. For example, when countries increase censorship during a particular period, negative online attitudes towards that country can surge, potentially leading to global conflicts.

In light of these risks, I strongly recommend that world leaders adopt the model shown in the ultimate game section of this paper and relinquish some control over their citizens' cyberspace behaviors. Allowing citizens to achieve cyberspace autonomy can benefit the economy by promoting innovation and entrepreneurship, stabilize the political system by reducing dissent and increasing trust in government, and contribute to a more peaceful future by reducing the risk of global conflicts. It is my hope that world leaders will recognize the importance of taking action to address this critical issue and work towards a better future for al

Appendix

Table 10. *Database of political party change*

Country name	frequency of change	change in the political party	year	start year	end year	total years of election
United States	0.115854	19	1857-2021	1857	2021	164
Canada	0.116883	18	1867-2021	1867	2021	154
China	0.009174	1	1912-2021	1912	2021	109
Germany	0.140625	9	1949-2013	1949	2013	64
Japan	0.034483	2	1955-2013	1955	2013	58
Australia	0.173554	21	1901-2022	1901	2022	121
United Kingdom	0.18894	41	1802-2019	1802	2019	217
Egypt	0.0625	6	1924-2020	1924	2020	96
Sudan	0	0	1956-2020	1956	2020	64
Belarus	0	0	1991-2020	1991	2020	29
Russia	0.1875	6	1990-2022	1990	2022	32
Estonia	0.151515	15	1920-2019	1920	2019	99
Cuba	0	0	1902-2020	1902	2020	118
Uganda	0.096774	6	1958-2020	1958	2020	62
India	0.088235	6	1951-2019	1951	2019	68
Indonesia	0.071429	3	1977-2019	1977	2019	42
Bangladesh	0.083333	4	1970-2018	1970	2018	48

Table 11. *Database of Reddit attitude*

Country_name	2022/03-2022/04	2022/04-2022/05	2022/05-2022/06	2022/06-2022/07	2022/07-2022/08
Pakistan	-3.96E-05	-0.00065	-0.00065	-0.00037	-0.00084
Norway	-1.32E-05	-0.00033	-0.00016	-5.27E-05	-0.00011
Turkey	-0.00015	-0.00065	-0.00097	-0.00126	-0.00123
Italy	-1.32E-05	-0.00114	-0.00057	-0.00042	-0.00078
Russia	-0.00194	-0.06622	-0.05349	-0.0513	-0.05431
China	-0.00066	-0.00627	-0.00465	-0.00311	-0.00407
Togo	-1.32E-05	-8.14E-05	-4.04E-05	-2.63E-05	-1.86E-05
Australia	-3.96E-05	-8.14E-05	-0.00061	-0.00047	-0.00034
Ukraine	-0.00096	-0.01002	-0.00764	-0.01302	-0.01378
Nigeria	-3.96E-05	-0.00033	-0.0002	-0.00016	-0.00024
Taiwan	-0.00022	-0.00212	-0.00137	-0.0009	-0.00095

Japan	0.000119	-0.00016	-0.00049	0.000211	-0.00013
France	-0.00013	-0.00261	-0.00129	-0.001	-0.00119
Hungary	7.93E-05	0.000977	0.000364	-0.00016	-0.00039
Poland	-7.93E-05	-0.00073	-0.00036	-0.00016	-0.0002
Kuwait	-1.32E-05	-8.14E-05	-0.00012	-7.90E-05	-3.73E-05
Qatar	-7.93E-05	-0.00041	-0.0004	-0.00026	-0.00019
Latvia	-3.96E-05	-0.00057	-0.00028	-0.00018	-0.00017
Spain	-0.00036	-0.00204	-0.00113	-0.00058	-0.00045
Indonesia	-6.61E-05	-0.00041	-0.0002	-0.00018	-0.0002
Bulgaria	-3.96E-05	-0.00033	-0.00032	-0.00016	-0.00015
Brazil	-3.96E-05	-0.00033	-0.00024	-0.00021	-0.00013
Thailand	-1.32E-05	-0.00024	-0.0004	-0.00016	-5.59E-05
Israel	-0.00052	-0.011	-0.0057	-0.00711	-0.0057
Ethiopia	-3.96E-05	-0.00033	-0.0004	-0.00032	-0.00024
Finland	-3.96E-05	8.14E-05	-0.0002	0.000395	0.000428
Iran	-0.0002	-0.0044	-0.00315	-0.00269	-0.00238
Switzerland	-3.96E-05	-0.00049	-0.00024	-0.00013	-9.31E-05
Congo	-1.32E-05	-0.00041	-0.00024	-0.00016	-0.00011
New Zealand	1.32E-05	8.14E-05	4.04E-05	2.63E-05	1.86E-05
Greece	-1.32E-05	-0.00016	-8.09E-05	0.000158	3.73E-05
Croatia	-1.32E-05	0.000244	0.000121	5.27E-05	1.86E-05
Singapore	-1.32E-05	-0.00016	-0.0002	-0.00016	-0.00022
Portugal	-3.96E-05	-0.00024	-8.09E-05	-5.27E-05	-3.73E-05
Cyprus	-1.32E-05	-8.14E-05	-8.09E-05	-5.27E-05	-1.86E-05
Estonia	-0.00011	-0.00244	-0.00133	-0.00095	-0.00078
Austria	-2.64E-05	-0.00016	-8.09E-05	-7.90E-05	-0.00013
Mexico	-6.61E-05	-0.00098	-0.00057	-0.00042	-0.00037
Mali	-2.64E-05	-0.00024	-0.00012	-0.00011	-0.00015
Afghanistan	-5.29E-05	-0.00041	-0.00024	-0.00018	-0.00015
Kenya	-1.32E-05	-8.14E-05	-4.04E-05	-7.90E-05	-9.31E-05
Honduras	1.32E-05	0.000163	8.09E-05	5.27E-05	3.73E-05
Guyana	-1.32E-05	-8.14E-05	-4.04E-05	-2.63E-05	-1.86E-05
Sweden	3.96E-05	0.000326	0.000162	0.000316	0.000335
Ireland	-1.32E-05	-0.00024	-4.04E-05	-0.00011	-9.31E-05
Armenia	2.64E-05	0.000244	0.000121	5.27E-05	0.00013
Belarus	2.64E-05	-0.00033	-0.00016	-0.00055	-0.00056
Chad	1.32E-05	8.14E-05	-4.04E-05	-2.63E-05	-1.86E-05
Azerbaijan	-3.96E-05	-0.00033	-0.00016	-0.00011	-9.31E-05
Iraq	-1.32E-05	-0.00057	-0.00032	-0.00034	-0.00024
Senegal	1.32E-05	8.14E-05	4.04E-05	2.63E-05	1.86E-05
Chile	-1.32E-05	-8.14E-05	-4.04E-05	-2.63E-05	-1.86E-05

Methodology for sentiment analysis

- Collect data from Reddit in r/worldnews
- Natural Language ToolKit(NLTK) analyzes the title of the Reddit post and returns the sentiment value for the sentence
- Assign a value for each of the words in the sentence based on the relationship between each word
  - Assign a higher value if it is described
  - Loop through the sentence

  - The highest value word will return as the subject
- Make a Numpy array that includes all the subjects and the attitude toward it
- Check if the subject is in the country array, and then add an attitude value to it
- Check if the subject is in the politician array, then add attitude value to the country array

Peaceful Model/War Model

- The peaceful model means that the countries cannot move around and start wars with each other; this is achieved by disabling the movements of the countries from the output of the neural network.
- Neural network specifications:
  - The numpy dot function is used for neural network simulations
  - The three outputs are x, y, and freedom
  - After each generation, there will be a mutation in the two layers of the neural network in terms of changing a specific gene that is randomly selected
  - Only countries with genes that make them survive the generation are passed on to the following simulation; thus, the neural network can learn strategies to survive the ultimate game.
- Simulation time: 54 minutes 36 seconds
- Simulation hardware: Quadro P620 graphics card, i7 tenth generation CPU
- Programming environment: Python on Pycharm editor(windows)
- Necessary packages for Python
  - Matplotlib
  - Numpy
  - Keras
  - Tensorflow > 1.15.0
  - Math
  - Seaborn
- The war model gives countries freedom to move around the simulation and start "wars" with other countries, as described in the Ultimate Game section of the paper

Short-term/Long-term simulation

- Short simulation – short-term simulation means that the country's cyberspace restrictions will not affect human development and resource use as people affected by the restrictions are still not mature enough to contribute to society yet. For example, children deprived of cyberspace usage will only have effects on the economy of a state when they are at the age of employment

- Long simulation – long-term simulation includes the effects of a country's cyberspace freedom restrictions in the calculation of their resource usa

References

Bonnaire, Céline, and Olivier Phan. "Relationships between Parental Attitudes, Family Functioning and Internet Gaming Disorder in Adolescents Attending School." *Psychiatry Research*, vol. 255, Sept. 2017, pp. 104–10. DOI.org (Crossref), https://doi.org/10.1016/j.psychres.2017.05.030.

China Population (2022) Live — Countrymeters. https://countrymeters.info/en/China.Accessed 4 Sept. 2022.

Cyberspace. link.springer.com, https://link.springer.com/book/10.1007/978-3-319-54975-0. Accessed 22 Aug. 2022.

Dreher, Axel. The Political Leaders' Affiliation Database (PLAD). Harvard Dataverse, 2020. DOI.org (Datacite), https://doi.org/10.7910/DVN/YUS575.

Egypt Jadaliyya. 16 Nov. 2011, https://web.archive.org/web/20111116034051/http://egypt.jadaliyya.com/.

FactsMaps. "PISA 2018 Worldwide Ranking - Average Score of Mathematics, Science and Reading." FactsMaps, 5 Dec. 2019, https://factsmaps.com/pisa-2018-worldwide-ranking-average-score-of-mathematics-science-reading/.

"Freedom House Index: Internet Freedom in Selected Countries 2021." Statista, https://www.statista.com/statistics/272533/degree-of-internet-freedom-in-selected-countries/. Accessed 23 Aug. 2022.

GDP (Current US$) | Data. https://data.worldbank.org/indicator/NY.GDP.MKTP.CD. Accessed 25 Aug. 2022.

Ghayoumi, Mehdi. Deep Learning in Practice. 1st ed., Chapman and Hall/CRC, 2021. DOI.org (Crossref), https://doi.org/10.1201/9781003025818.

Growth in GDP per Capita, Productivity and ULC. https://stats.oecd.org/Index.aspx?DataSetCode=PDB_GR. Accessed 23 Aug. 2022.

Hill, Brittany. "Parents Perceptions of the Internet and Its Effects on Their Children." Honors Theses, May 2017, https://scholar.utc.edu/honors-theses/98.

"History of Elections in Uganda." Electoral Commission, 29 Dec. 2015, https://www.ec.or.ug/info/history-elctions-uganda.

Mainwaring, Sarah. "Always in Control? Sovereign States in Cyberspace." *European Journal of International Security*, vol. 5, no. 2, June 2020, pp. 215–32. DOI.org (Crossref), https://doi.org/10.1017/eis.2020.4.

MSD. Censorship In New Zealand: The Policy Challenges Of New Technology - Ministry of Social Development. MSD. www.msd.govt.nz, https://www.msd.govt.nz/about-msd-and-our-work/publications-resources/journals-and-magazines/social-policy-journal/spj19/censorship-new-zealand-challenges19-pages1-13.html. Accessed 4 Sept. 2022.

Mayer, Chloe. "Russian Oligarch Sergey Protosenya and Family Found Dead in Spain." *Newsweek*, 21 Apr. 2022, https://www.newsweek.com/russian-oligarch-family-dead-spain-1699660.

Nohlen, Dieter, and Philip Stöver, editors. Elections in Europe: A Data Handbook. 1. Ed, Nomos, 2010.

Perez, Yael Valerie, and Yahel Ben-David. "Internet as Freedom – Does the Internet Enhance the Freedoms People Enjoy?" *Information Technology for Development*, vol. 18, no. 4, Oct. 2012, pp. 293–310. DOI.org (Crossref), https://doi.org/10.1080/02681102.2011.643203.

Phishing Attacks: Defending Your Organisation. https://www.ncsc.gov.uk/guidance/phishing. Accessed 27 Aug. 2022.

"Phishing: Distribution of Attacks by Country 2021." Statista, https://www.statista.com/statistics/266362/phishing-attacks-country/. Accessed 27 Aug. 2022.

PISA. https://www.oecd.org/pisa/. Accessed 27 Aug. 2022.

"Political Stability by Country, around the World." TheGlobalEconomy.Com, https://www.theglobaleconomy.com/rankings/wb_political_stability/. Accessed 23 Aug. 2022.

"Global_economy" TheGlobalEconomy.Com, https://www.theglobaleconomy.com, rankings/wb_political_stability/. Accessed 30 Aug. 2022.

Psephos - Adam Carr's Election Archive. http://psephos.adam-carr.net/countries/c/colombia/colombiamapsindex.shtml. Accessed 27 Aug. 2022.

Russia Population (2022) - Worldometer. https://www.worldometers.info/world-population/russia-population/#:~:text=The%20current%20population%20of%20the,the%20latest%20United%20Nations%20data. Accessed 4 Sept. 2022.

"Russia-Ukraine War: Moscow Politician Gets 7 Years for Denouncing War." BBC News, 8 July 2022. www.bbc.co.uk, https://www.bbc.com/news/world-europe-62092196.

Rychlík, Jan. "The 'Velvet Split ' of Czechoslovakia (1989-1992)." *Politeja*, vol. 15, no. 6(57), Aug. 2019, pp. 169–87. DOI.org (Crossref), https://doi.org/10.12797/Politeja.15.2018.57.10.

---. "The 'Velvet Split' of Czechoslovakia (1989-1992)." *Politeja*, vol. 15, no. 6(57), Aug. 2019, pp. 169–87. DOI.org (Crossref), https://doi.org/10.12797/Politeja.15.2018.57.10.

Sinpeng, Aim. "State Repression in Cyberspace: The Case of Thailand: State Repression in Cyberspace." *Asian Politics & Policy*, vol. 5, no. 3, July 2013, pp. 421–40. DOI.org (Crossref), https://doi.org/10.1111/aspp.12036.

---. "State Repression in Cyberspace: The Case of Thailand: State Repression in Cyberspace." *Asian Politics & Policy*, vol. 5, no. 3, July 2013, pp. 421–40. DOI.org (Crossref), https://doi.org/10.1111/aspp.12036.

Sumo3000. "Top 20 Countries Found to Have the Most Cybercrime." Remove Spyware & Malware with SpyHunter - EnigmaSoft Ltd, 9 July 2009, https://www.enigmasoftware.com/top-20-countries-the-most-cybercrime/.

Ten, Alexandr, et al. "Intrinsic Rewards in Human Curiosity-Driven Exploration: An Empirical Study." Proceedings of the Annual Meeting of the Cognitive Science Society, vol. 43, no. 43, 2021. escholarship.org, https:// escholarship.org/uc/item/13b6p5ms.

Troianovski, Anton, and Valeriya Safronova. "Russia Takes Censorship to New Extremes, Stifling War Coverage." *The New York Times*, 4 Mar. 2022. NYTimes.com,https://www.nytimes.com/2022/03/04/world/europe/russia- censorship-media-crackdown.html.

Zhang, Han. "The Censorship Machine Erasing China's Feminist Movement." *The New Yorker*, 29 Aug. 2022. www.newyorker.com, https://www.newyorker.com/news/news-desk/the-censorship-machine-erasing-chinas-feminist-movement.

Buckley, Chris. "China Tightens Limits for Young Online Gamers and Bans School Night Play." *The New York Times*, 30 Aug. 2021. NYTimes.com, https://www.nytimes.com/2021/08/30/business/media/china-online-games.html.

A Discourse on Utopia and the New World: Political Models in *The Tempest*

Lingchen Wang

Author Background: *Lingchen Wang grew up in China and currently attends Shanghai Starriver Bilingual School in Shanghai, China. His Pioneer research concentration was in the field of literature and titled "The Power of Shakespeare."*

1. Introduction

The Renaissance and Age of Discovery during the Early Modern period signaled the beginning of a political exploration—a search for new forms of government. The discovery of the Americas and the Caribbean in the fifteenth and sixteenth century revealed societies with radically different political structures. In the diaries and records of the travelers, European political thinkers saw the possibility that the Golden Age myth of Ovid, an early tale of romanticized ideal society, was not a historical period of remote antiquity, but a contemporary reality (Wells 29). As they juxtaposed the newfound political structures with their own, they began envisioning the ideal form of government and contemplating whether it can be applied to the newly discovered societies.

William Shakespeare was commonly considered to be among the pre-eminent Renaissance political thinkers, though at the same time dissimilar to them in the sense that he has not left any treatises or texts that systematically examine governmental principles or actions. Nevertheless, he was undeniably engaged with politics and governmental affairs, though as a "knowledgeable, insightful person" who, as a participant in public theater, had to deal with censorship authorities (Frazer, "Shakespeare's Politics" 506). Due to these circumstances, Shakespeare's political insights are often presented through allusions in his plays, sometimes in the form of character commentary. As a result, his works often lend themselves to political interpretations. This is evident in his last play, *The Tempest*, which takes place on a Mediterranean island situated somewhere between Naples and Tunis. It opens with the arrival of King Alonso and his entourage after surviving a shipwreck and ends with their merry departure accompanied by Prospero, the former Duke of Milan who spent years on the island in exile, and Miranda, his daughter.

The play includes multiple references to the New World. The spirit Ariel reports to Prospero, for example, that he fetched dew from the "still-vexed Bermoothes" (1.2.271-2), suggesting the play's concern with Jacobean colonial projects in the New World (Stanivukovic 91). Caliban mentions that he and his mother Sycorax worship a god named "Setebos" (1.2.449), which is, in reality, a god worshiped by South American Natives (Frey 1). A more direct reference is when Miranda, after meeting the courtiers and witnessing their finery, exclaims:

"O brave new world / That has such people in 't!" (5.1.217).

All of these signs suggest the presence of the New World imagery in the play, especially in the context of materials and accounts of travel considered readily available to Shakespeare. The most prominent one is, perhaps, William Strachey's account of the shipwreck of the flagship *Sea Venture*.[1] After the ship was blown off its course by a hurricane and wrecked off the coast of Bermuda, the passengers initiated England's colonization of Bermuda and, after a year, built two vessels with which they sailed to the Jamestown Colony, their original destination. These New World materials began the topical discussion of colonialism and the concept of the "brave new world." Many Shakespearean scholars argue that these materials and accounts of travel cast influence on the playwright as he composed the play.[2] Edmond Malone argues that Shakespeare bases the play partially on the accounts of the storm and shipwreck experienced by Sir Thomas Gates (who was a passenger aboard the *Sea Venture*).[3] Based on these contextual assertions, Prospero's island in the play, as a remote and scarcely populated natural land, is an ideal place for political hypotheses and experimentations.

On his island in *The Tempest*, Shakespeare investigates utopian concepts in relation to power, society, and politics. Through the characters Gonzalo and Prospero, Shakespeare presents different political models with respect to utopianism. Gonzalo, in his vision of how the island should be governed, remains hypothetical and even satiric, while Prospero's vision is more fully realized. Gonzalo's vision directly alludes to French Renaissance philosopher Michel de Montaigne's *Of the Cannibals* and Ovid's myth of the Golden Age. His idealistic stance is caustically interrupted by Sabastian and Antonio; he himself is later disillusioned by a series of attempted violent usurpations. Prospero, whose name suggests evolution in Italian, leaps between different political models. He develops from a failed humanist ruler to a ruler who implements Machiavellian politics to create a society that fundamentally stems from Sir Thomas More's *Utopia*. His use of magic to exert control and exploit others is absolutist and monarchical, but his politics take an abrupt turn to Christianity at the denouement of the play as he forgives the usurpers and renounces his magic. Shakespeare's use of botanical metaphors in his language reveals the dialogic nature of the characters' arguments and assists in the staging of conflicts between the models. Ultimately, the play is not a response to the utopian discourse, but an experimentation and rejection of three possible ideal political structures raised by three political philosophers. As I claim below, the

---

[1] For the complete narrative, see Strachey, *A True Reportory of the Wreck and Redemption of Sir Thomas Gates, Knight* in *A Voyage to Virginia in 1609*, ed. Louis B. Wright (Charlottesville, VA: Univ. Press of Virginia, 1965).

[2] *Narrative and Dramatic Sources of Shakespeare*, ed. Geoffrey Bullough (London: Routledge and Paul, 1975). VIII, 245. In addition, Sir Frank Kermode argues that when writing the play, Shakespeare has in mind certain Bermuda pamphlets and accounts of travel literature. Hallett Smith argues that Shakespeare's imagination would appear to have been stimulated by the exploration of the new world in his *Shakespeare's Romances: A Study of Some Ways of the Imagination* (San Marino: Huntington Library, 1972), p. 143.

[3] *Account of the Incidents from Which the Title and Part of the Story of Shakespeare's Tempest Were Derived* (London, 1808). Malone points out the connection between the play and Sir Thomas Gates, along with other Jamestown adventurers.

island itself is not a secluded utopia, but a representation of the disenchantment with utopianism. In the end, the political models all prove to be fragile and not a single one prevails over the others.

2. More, Montaigne, and Machiavelli's Models

Before situating the play in its political context, it is necessary to first examine the three principal political models proposed by three influential contemporary political philosophers—Thomas More, Michel de Montaigne, and Niccolò Machiavelli. The three models are all present in *The Tempest*'s discourse but displayed to different extents. Montaigne's primitivist argument in *Of the Cannibals* is presented directly as Gonzalo quotes from it in his speech. Machiavelli's argument in *The Prince* gained much popularity and has a deep connection with Shakespeare's writing, for they both focus on similar issues (such as how political power is maintained and how it connects to Christian morality) (Grady 123). More's argument in *Utopia* is implicitly connected to Shakespeare, but several characteristics of Prospero's island are relatable to More's conception of Utopian society.

Utopias are designed to posit better alternatives to existing social, political, and cultural systems (Bulger 1). The term is first coined by More in his book *Utopia*, in which readers are invited to the dialogue between More and the traveler Raphael Hythloday. Hythloday, under More's request, provides detailed descriptions of the political system of the island Utopia. The story itself is "a drama of More's mind;" he measures earthly states against a divine standard (Sanderlin 76). The standards of Utopian society are radically different from conventional European political models. For instance, Hythloday claims that private property must be entirely abolished for there to be just distribution of goods and for mortals to conduct business happily (More 38). The collective ownership and equal allocation of all resources echoes Plato's claim in *The Republic*.[4] Moreover, the Utopians have a collective disdain toward gold, which they consider a useless commodity that has no value in and of itself. Similarly, they recognize the inanity of any man "who considers himself a nobler fellow because he wears clothing of specially fine wool" (More 63). Shakespeare demonstrates this notion in *The Tempest* with the characters Stephano and Trinculo. As they arrive at the entrance of Prospero's lodging, they are drawn to the glistering apparel (4.1.245-55). Ironically, they favor a false display of social superiority when they have the chance to murder Prospero and become the real rulers of the island.

Another political model that challenges conventional European perceptions is that of Montaigne. In his essay *Of the Cannibals*, he gives detailed ethnographic descriptions of the traditions and ceremonies of the Tupinambá people in Brazil. In his portrayal, he mentions that there is "no metals, no use of wine or corn" (Montaigne 233). The same phrases are repeated by Gonzalo, which indicates that, whether he shares it or not, Shakespeare is consciously

[4] Cf. Plato, *The Republic* V: "In the first place, none must possess any private property save the indispensable. Secondly, none must have any habitation or treasure-house which is not open for all to enter at will" (311).

showing the readers Montaigne's view on human nature. Montaigne sharply challenges the idea that Europeans are more civilized, and hence superior, when compared to the "barbarous" indigenous people. He argues that "every man calls barbarous anything he is not accustomed to;" it is in this manner that they bastardize what is virtuous and natural by "adapting them to [their] own corrupt tastes" (Montaigne 231-32). Montaigne refutes the assumption that the European way should be viewed as the standard for civilized society. On the contrary, he inverts the process and argues that the so-called barbarians are civilized in a purer way, because they are produced "by Nature" and are not corrupted by greed and vice. Montaigne writes that the society he observes surpasses the descriptions of the "Age of Gold" (Montaigne 231). He even addresses Plato directly, using a series of negations as a rejection of fundamental elements of European society (similar to Gonzalo's commonwealth speech, which will be analyzed later in the essay):

> I would tell Plato that those people have no trade of any kind, no acquaintance with writing, no knowledge of numbers, no terms for governor or political superior, no practice of subordination or of riches or poverty, no contracts, no inheritances, no divided estates, no occupation but leisure... (Montaigne 233)

Montaigne reduces his model to the most primitive state without political interference and argues that the initial uncontaminated purity is a sign of civilization. Although Montaigne and More's political standpoints are both mirrored in the play (which will be discussed in the following sections), the two structures are different in nature. The Utopian qualities of More's structure need to be instituted, maintained, and upheld, whereas those of Montaigne's are intrinsic to the natural state and are worthy of appreciation. Prospero and Gonzalo display this difference in a subtle manner; Prospero's governance consists of active action, while Gonzalo's hypothesis focuses more on reduction. Gonzalo adheres to the primitivist principle and eliminates artificial elements of society to return it to its "original state of nature" (Montaigne 232).

The sense of disenchantment and fragility is inherent in both works. More invented the term "utopia" by combining the Greek adverb *ou*—"not"—with the noun *topos*—"place"—resulting in the translation "Noplace." However, the word also puns with *eutopia*, suggesting a happy or fortunate place. The appearance of happiness of More's Utopia disguises its unfounded nature. Hythloday, who is the narrator of the perfect society, has the name with the Greek meaning "nonsense peddler." When More suggests to Hythloday that he should enter the king's service, Hythloday firmly refuses, because he believes the "seeds of evil and corruption" in the king's mind will not allow him to accept wise advice (28). More himself, after hearing Hythloday's account of the Utopian commonwealth, admits that there are many features that he would "wish rather than expect to see" in his own European society (107). Montaigne, on the other hand, does not directly state the applicability of his political stance, but he gives the essay an ironic ending that seems jocular and concludes the sense of disillusionment that the New World model can be accepted by Europeans: "Not at all bad, that. – Ah! But they wear no breeches..." (Montaigne 241).

The third political model is the Machiavellian model in *The Prince*;

although it does not appear directly, it brought a paradigmatic shift to the way power is conceived and provides historical context for the play. Machiavelli focuses on the pragmatic utilization of power in politics. His insights into these subjects, which Arlene Oseman describes as "dispassionate and almost scientific," provide a new way to examine the political intrigue in Europe (8). Machiavelli concedes the "fallen" human nature and, in doing so, refuses the Christian hope of perfection that would lead to a divine structure of balance and justice (Oseman 9). He states that it is necessary for a prince to learn to "not be good" according to the circumstance, because "a man who wants to make a profession of good...must come to ruin among so many who are not good" (Machiavelli 61). A successful prince, in Machiavelli's political argument, is not always the hereditary or designated ruler, but the one who masters political maneuvering and utilizes it to acquire or maintain power. In the sixteenth century, Machiavelli was widely read and his importance as a political thinker was commonly acknowledged (Wells 27); although it is arguable whether or not Shakespeare and Machiavelli are directly correlated, the power-plays in Shakespeare's plays can be examined through Machiavelli's arguments.

3. Gonzalo and the Rejection of Primitivism

Shakespeare, through Gonzalo's political insights, seems to be initiating a satiric response to the primitivist approach to imagining the island. After his arrival, Gonzalo delivers a speech with allusive reference to Michel de Montaigne regarding the island as a theoretical commonwealth. His speech is crucial to the play's political discourse because of its direct quote from Montaigne's essay *Of the Cannibals*. Gonzalo's political postulation derives from Montaigne's, which is juxtaposed with Prospero's later in the play. His vision also draws from the myth of the Golden Age.

In his speech, Gonzalo details a specific list of requirements under his imaginary reign:

> I' th' commonwealth I would by contraries
> Execute all things, for no kind of traffic
> Would I admit; no name of magistrate;
> Letters should not be known; riches, poverty,
> And use of service, none; contract, succession,
> Bourn, bound of land, tilth, vineyard, none;
> No use of metal, corn, or wine, or oil;
> No occupation; all men idle, all,
> And women too, but innocent and pure;
> No sovereignty— (2.1.162-71)

Gonzalo's speech comes from a specific part of Montaigne's essay, where Montaigne elaborates on the purity and simplicity of the indigenous people by listing his rejection of the common elements of a Renaissance social structure. Montaigne's utopian vision (and Gonzalo's as well) is predicated on its closeness with "the original state of nature;" he romanticizes and civilizes

what is commonly considered primitive and barbaric, claiming that this is "Man's blessed early state" (232). With his view in mind, the origin of Gonzalo's assumption and his motivation can be revealed. As he arrives on an island with natural landscape uncontaminated by what Montaigne calls corrupt adaptations, Gonzalo is immediately inspired and recognizes the island's potential to become the primitive paradise Montaigne paints in his works.

Gonzalo's imagination is so noble and romantic that it seems unreal and absurd. The flaws of his attempt to employ Montaigne's primitivism and to realize the myth of Astraea are exposed almost immediately after his delivery of the famous commonwealth speech. Gonzalo asserts that he would rule with "no sovereignty" (2.1.171), governing with such perfection that it would "excel the Golden Age" (2.1.184). Sabastian and Antonio immediately underscore the fact that Gonzalo would be the king, and that "the latter end of his commonwealth forgets the beginning" (2.1.173-74). This is more than a disparaging remark; it reveals the underlying contradiction in Gonzalo's political structure. To be the king of a commonwealth that should not be subject to any form of sovereignty in the first place is ironic. The necessity of a monarch to institute a political regime that is republican in nature is intrinsically paradoxical. After hearing the speech, Alonso requests that Gonzalo cease his talking. He then refers to the entire political hypothesis as "nothing," a word that is repeated by Sabastian and Antonio as they heap scorn on Gonzalo (2.1.188). Alonso's reply as a ruler shows that what Gonzalo proposes is a static model of perfection that is unrealizable anywhere (Bulger 4). The word creates a gap between the vision's idealism and its feasibility. Gonzalo then calls his own theory "nothing," because Antonio and Sabastian fail to realize that Gonzalo is providing an invitation to open a debate (Laghi 189). Gonzalo's primitivism is not naïve, as Laghi argues, but is so overly-broad and idealistic that it eventually amounts to nothing. Through the three characters' collective negative reaction to Gonzalo's speech, Shakespeare unmasks the unrealistic condition to abolish sovereignty.

Furthermore, when he delivers the commonwealth speech, Gonzalo is not yet aware that he is overlooking human nature's proclivity for corruption and selfishness. As the play progresses, two attempted usurpations take place. The first is Sebastian and Antonio's attempt to murder the King of Naples along with his courtiers in their sleep. The second is Caliban, Stephano, and Trinculo's attempt to assume control of the island by killing Prospero. The two usurpations are committed by characters representing two different classes of the social hierarchy, the aristocrat and the plebeian. The two separated groups of characters both display a tendency toward criminal and violent acts to overthrow their political superiors. Intriguingly, through their violent schemes, they are adhering to the Machiavellian principle to murder the powerful and commit injuries all at a stroke (Machiavelli 38). Antonio and Sebastian's plan to supplant Alonso resembles what was done to Prospero in Milan. Caliban's decision to accept Stephano as his new master in order to murder his previous master is also a pragmatic exploitation of power. This can be read as a sign of how Machiavelli's realistic political concepts and his doubtful attitude about human nature could have impacted Shakespeare and other serious writers of sixteenth century Europe. Even on an isolated island where the primary goal is no longer realizing political aspirations but simply surviving, the characters are still individually motivated by their lust for power. It seems to demonstrate in

action the innate evil of human nature that Protestant theology and Machiavellian empiricism assert (Wells 28).

4. Prospero's Politics and Investigation

Initially, Shakespeare shapes Prospero as the representation of humanism, a ruler who dedicates his time to the study of liberal arts. His reign was weak and passive, leaving him exposed to the risk of potential usurpations. At the beginning of the play, Prospero reveals to Miranda how he lost his dukedom:

> I, thus neglecting worldly ends, all dedicated
> To closeness and the bettering of my mind
> With that which, but by being so retired,
> O'erprized all popular rate, in my false brother
> Awaked an evil nature, and my trust
> Like a good parent, did beget of him
> A falsehood in its contrary as great
> As my trust was, which had indeed no limit,
> A confidence sans bound. (1.2.109-17)

Prospero resembles the figure of a humanist scholar instead of a realistic ruler, prioritizing reading and "bettering of [his] mind" over his political ambitions. He disregards his duties as a duke and fails to fulfill the task of governing. His abuse of power through negligence and his consequent failure to control his dukedom reflects the "discontent" of humanism, or specifically, the humanist political pragmatism (Stanivukovic 96). Prospero's devotion to his humanist studies was not translated into his ability to institute an effective government.

The significance of Prospero's account of the past exceeds a mere recollection of the past narrated by a father to his daughter; it lies in his acknowledgment of his past mistake. Prospero recognizes that he failed to balance his passion for liberal arts and his governing. Moreover, he does not simply condemn Antonio for his exile, but admits his own imprudent and misplaced trust in Antonio in the first place. Prospero's reconstruction of the past reveals his awareness of the flaws during his reign. It also reflects his determination to seek a revised form of government and institute it on the island. His exile becomes a second opportunity for him to learn from previous failure and reimagine his governmental structure.

Interestingly, the process Prospero takes to reform the island seems noticeably antithetical to More's description of the Utopians. According to Raphael Hythloday's narration, the Utopians learned every art of the Roman empire after a shipwreck brought several Romans and Egyptians ashore (More 39). In other words, the perfect social structure of Utopia is inspired by ancient and existing models; the inhabitants of the island learned from the interlopers, and the interlopers soon became part of Utopia and never departed. In *The Tempest,* the process is reversed when Prospero assumes full control of the island and shapes it according to his own vision. His negligence of the preexisting primitive system on the island is opposed to Montaigne's political

view that Gonzalo represents.

Although Prospero resumes his political regime on the island, the nature of his reign is altered with the development of his character as a ruler. He "exchanges republican rule as a duke in Milan for authoritarian and absolutist rule," thereby becoming a ruler who utilizes available resources to enhance his domination (Frazer, "Political Power" 363). Prospero's politics is composed of two necessary parts: his use of magic as a political resource and his control of other characters. Both manifest the signs of extreme absolutism and monarchism; Prospero becomes calculating, exploitative, and controlling. This is opposed to Montaigne and More's political claims, but it is inclined toward Machiavelli's. However, Prospero's method takes an abrupt Christian twist at the end, rejecting the Machiavellian model and further establishing him as a political experimenter. Shakespeare seems to be staging a conflict between the models as Prospero adopts different ways of governance.

When Prospero calls upon the spirits to host a feast for King Alonso and his party, he is capable of being invisible as he observes them (3.3.22 SD). Prospero, at the end of the play, claims that he has "bedimmed the noontide sun" and "called forth the mutinous winds" by his "potent art" (5.1.50-9). Magic gives Prospero the ability to control plot development; he can produce a tempest to bring the characters to his island and charm them to sleep as he likes. Magic is the manifestation of his omnipotence within the domain of the island. In Elizabethan and Jacobean Britain, "magic loomed large" as a factor of the everyday, in state rule, and for sovereign concerns (Frazer, "Political Power" 360). Frazer's argument establishes magic as the proper pursuit of monarchs, yet I argue that Prospero's magic is not a subject of his pursuit, but rather a means to achieve certain political ends. After he arrives on the island, Prospero enslaves Caliban, who admits in an aside: "I must obey. His art is of such power / It would control my dam's god, Setebos, / And make a vassal of him" (1.2.448-50). Prospero also keeps the spirit Ariel in bondage, constantly reminding him of the freedom that he will eventually grant him. Through keeping the two original inhabitants of the island under his control, Prospero exploits Caliban's physical labor and Ariel's spiritual magic. Although not as noticeable, Prospero's control over Miranda also reveals underlying political purposes. Miranda is educated to be the silent, subservient "foot" to Prospero's head (1.2.569). Her encounter with Ferdinand and their later marriage can largely be attributed to Prospero's careful arrangement and use of magic.

Prospero achieves two purposes with the politics he establishes and his absolutist control. The first goal is the restoration of his position as the Duke of Milan. After acknowledging the penitence of the usurpers, Prospero shows the mercy of a humanist ruler and forgives them on condition that his dukedom is reinstated (5.1.36). The second goal is to secure a lasting alliance between Milan and Naples, an anticipated outcome of his daughter's marriage with Ferdinand (Frazer, "Political Power" 363). Both goals are achieved with the use of either his magic or his manipulation of others. Shakespeare depicts magic as a political shortcut, a substitute for Machiavellian scheming and military exercise (Oseman 13). The use of magic to achieve political goals is less complex than Machiavellian political maneuvering and more illustrative of Prospero's omnipotence.

The island Prospero shapes with his magic seems to share the set of

rules that governs Utopia. Although Prospero's island does not consist of fifty-four cities and is not as populated as Utopia, he rules the island in accordance to some basic utopian principles. As the traveler Raphael Hythloday recounts, all goods are shared and equally allocated in Utopia, which he argues is the only path to public welfare (More 37). On Prospero's island, natural resources are abundant and shared among the members of the society after Caliban discloses to him "all the qualities o' th' isle" (1.2.403). He also demonstrates the appreciation of knowledge (*studia humanitatis*) that Hythloday stresses is important in Utopia; Prospero praises Gonzalo for furnishing him from his library with volumes that he "prizes over [his] dukedom" (1.2.198-199). His education of Miranda is a reflection of the ideals of Utopian education:

> Here in this island we arrived, and here
> Have I, thy schoolmaster, made thee more profit
> Than other princes can, that have more time
> For vainer hours and tutors not so careful. (1.2.205-8)

Prospero's education proves successful in Miranda, but not in Caliban. Hythloday, when describing slavery in Utopia, underscores the fact that it is non-hereditary in nature (More 77). Prospero's treatment of Caliban adheres to this principle; although he is the son of Sycorax, Prospero provides him with proper education and attempts to adapt him to civilization. To this, Caliban replies:

> You taught me language, and my profit on 't
> Is I know how to curse. The red plague rid you
> For learning me your language! (1.2.437-9)

Caliban's resistance is a sign of the island's deviation from utopian ideals. After Prospero frees Ariel and accepts Caliban (by educating him and living with him), his rule approaches Utopia because of the island's abundant resources and his use of magic to assume complete political control. However, after Caliban's attempted rape of Miranda and his subsequent condemnation to slavery, the island deviates from the perfect society to become a "reduced scale model of the society" from which Prospero was exiled (Laghi 188).

Prospero, at the end of the play, acknowledges the flaws of his political structure. He calls his own art "rough magic" and renounces further use of it (5.1.59-66). The art that was "potent" is suddenly regarded as "rough." Critics fill in the subtextual argument in two ways. The word "rough" could either imply Prospero's strong sense of disgust toward the magic and himself or simply mean "unrefined," thereby signaling a metamorphosis of his political project (Corfield 32). Whichever is the case, abjuring the magic that has functioned as his political tool to implement his Machiavellian reign indicates a sharp rejection of Machiavellian politics. This argument is fortified by Prospero's final decision to forgive the usurpers. In contrast to the political manipulations and calculations of Antonio, Sebastian, Stephano, Trinculo, and Caliban, Prospero's solution at the end is permeated with the Christian notion of forgiveness and reconciliation (Oseman 13). Such Christian virtues were rejected by Machiavelli, a foe to Christianity who unambiguously instructed princes to resort to evil and use it

well in order to maintain power (Machiavelli 37). Hence, Prospero's choice to forgive signals a deviation from Machiavelli's model and marks a clear volte-face of his political project. His transformation from the anti-Machiavellian Duke of Milan to a Machiavellian ruler of the island, and to a morally superior Christian ruler at the end, mirrors an exploration of Machiavelli's political principles and the eventual rejection of it.

Furthermore, Christianity in this context functions not only as a rejection, but also as a possible alternative that could further the construction of the ideal society. Prospero's choice reveals an underlying assertion that an ideal society can only come with an elevation of human nature; this elevation is not brought by books and human knowledge, but by an exercise of Christian virtues (Ebner 161). Despite this, Shakespeare's denouement again seems to reveal himself imbued with a sense of disillusionment. Stephano, Gonzalo, and Alonso immediately repented their wrongdoing; even Caliban promises to "be wise hereafter / and seek for grace" (5.1.351-2). However, Antonia and Sabastian remain ominously silent, suggesting that the political tension is yet to be resolved (Oseman 17). This unresolved tension implies a possibility of political rebellion and will prevent further construction of a utopian society.

Ultimately, Prospero chooses to depart from the island and return to Milan as the duke. His act of forsaking his utopian experiment on the island indicates his realization that a perfect political structure cannot be achieved on the island with his magic. The sense of disillusionment signals a new beginning as Prospero embarks on his renewed political quest back in Milan as a refined ruler.

5. Utopian Dialogue

To demonstrate that the political models presented by the two characters are dialogical rather than fixed, Shakespeare uses botanical metaphors that connect Prospero and Gonzalo's thoughts. These metaphors suggest that the two characters are not separate, but are both active participants of the play's utopian discourse. After his arrival, Gonzalo starts his conversation with the exclamations that on this island there is "everything advantageous to life" (2.1.51-2) and that the grass is "lush and lusty" (2.1.55-6). Gonzalo's use of language points to the imagery of a Renaissance garden, which, in political discourse, is often invoked as a metaphor for the ideal republic (Samson 6). Shakespeare's reference to horticultural imagery is shown through the mockery of Sabastian and Antonio. They remark that Gonzalo will "carry this island home in his pocket" and, "sowing the kernels of it in the sea, bring forth more islands," to which Gonzalo responds "ay" (2.1.94-8). The two characters fail to realize the significance of their own ridicule; at this point, Gonzalo seems to be pondering over the island as a garden of political ideals and how it may be applied to the larger garden of Europe. Gonzalo considers the island to be a sample of ideal society that can be applied to the European political structure.

Prospero's use of horticultural metaphors appears even earlier than Gonzalo's. When narrating how he was overthrown by his brother Antonio, Prospero calls Antonio "the ivy which had hid [Prospero's] princely trunk and

sucked [Prospero's] verdure out on 't" (1.2.104-6). The ivy slowly sucks away the plant's vitality, similar to how Antonio gradually takes power away from Prospero and rises to become the Duke of Milan. Prospero uses the phrase "to trash" when describing Antonio's usurpation (1.2.100). In Elizabethan gardening books, "trash," as a verb, means "to cut away superfluities" (Laghi 183). The underlying significance of the word constructs a botanical image, in which Antonio is like the gardener discarding Prospero after acquiring more and more control of the dukedom. Intriguingly, the art of gardening is compared to the art of governing; the former manipulates nature to help seeds flourish, while the latter forges a suitable legal system to create prosperity (Laghi 183). For Prospero, the island is the metaphoric garden that represents the creation and enforcement of law; since he was exiled from his Edenic garden (which is Milan), he is forced to become "the maker of an alternate space with a new legal system" (Carpi 39). This metaphor deepens Prospero's utopian experimentation and reinforces his role as a humanist creator.

The common use of horticultural language in Gonzalo and Prospero's speech indicates that they are engaging in a political dialogue. Their ideas epitomize different political stances, and the island is, consequently, the center of the clash of different political structures. Their politics, when considered collectively, reveal Shakespeare's disillusionment with the utopian models.

6. Conclusion

The Age of Discovery and the travelers' accounts of the New World inspired Renaissance political thinkers to imagine the ideal form of government and the possibility of applying it to the newly discovered lands. It seems more than likely that Shakespeare was aware of this as he composed his final play, *The Tempest*; many argue that the shipwreck in the play is based on the historical travel of Sir Thomas Gates. The correspondence between the play and the shipwreck of the *Sea Venture* en route to Jamestown is an explicit indication of the play's political subtext, which is clearly recognized by literary critics as a path to a variety of political interpretations. Thus, the island becomes the ideal platform on which Shakespeare stages the conflicts between More, Montaigne, and Machiavelli's political models.

Gonzalo's political vision, quoted directly from Montaigne, represents the primitivist view of the New World. Through the king's courtiers' caustic refusal and the two attempted usurpations, Gonzalo's stance proves to be satiric and overly idealistic. Prospero, on the other hand, experiments with different models throughout the play. Initially, his inattentive humanist reign was supplanted by his brother's scheming. On the island, he learns from his failure and utilizes magic to become a Machiavellian ruler, constructing a political system that intrinsically epitomizes that of More's Utopia. During his rule, he reveals himself to be an authoritarian and absolutist ruler who is cunning and controlling. However, his revenge scheme takes an abrupt deviation at the end; Prospero embraces the Christian virtue of mercy and forgives the usurpers.

The significance of this interpretation lies in the juxtaposition and interplay of three renowned political models and their overall contribution to

the play's utopian discourse. Although Shakespeare provides a positive ending with the characters' reconciliation, Prospero's restored dukedom, and the marriage of Miranda and Ferdinand, the final scene is suffused with disillusionment. None of the three models involved in the experimentation emerges superior over the others; yet despite the rejection of these models, the play's political exploration is not at an end. Prospero's renewed perception is shown by his humble request for freedom and forgiveness: "Now my charms are all o'erthrown, / and what strength I have 's mine own, / which is most faint...As you from crimes would pardoned be, / Let your indulgence set me free" (Epilogue.1-3, 19-20). Thus, the discourse extends beyond the play to form an interaction between Prospero and the audience. Shakespeare implies that the utopian models are intrinsically brittle and unfeasible, but the utopian discourse is boundless in nature. It fostered further exploration of related topics and inspired many later writers, most notably Aldous Huxley and his dystopian novel *Brave New World*, the title of which derives from Miranda's speech.

Paralleling the perpetual nature of Shakespeare's political discourse, Prospero's exploration does not cease at the denouement; rather, it is taken to a new stage after its metamorphosis. After renunciation of his magic and island, Prospero will continue his search for a new and ideal form of government as the restored Duke of Milan, his political quest expanding from the microcosm to the social macrocosm.

Works Cited

Bulger, Thomas. "The Utopic Structure of *The Tempest*." *Utopian Studies*, vol. 5, no. 1, 1994, pp. 38–47. JSTOR, http://www.jstor.org/stable/20719247. Accessed 16 Aug. 2022.

Carpi, Daniela. "The Garden as the Law in the Renaissance: A Nature Metaphor in a Legal Setting." *Pólemos*, vol. 6, no. 1, 2012, pp. 33–48. De Gruyter, https://doi.org/10.1515/pol-2012-0003. Accessed 16 Aug. 2022.

Corfield, Cosmo. "Why Does Prospero Abjure His 'Rough Magic'?" *Shakespeare Quarterly*, vol. 36, no. 1, 1985, pp. 31–48. JSTOR, https://doi.org/10.2307/2870079. Accessed 16 Jan. 2023.

Ebner, Dean. "*The Tempest*: Rebellion and the Ideal State." *Shakespeare Quarterly*, vol. 16, no. 2, 1965, pp. 161–73. JSTOR, https://doi.org/10.2307/2868262. Accessed 16 Jan. 2023.

Frazer, Elizabeth. "Political Power and Magic." *Journal of Political Power*, vol. 11, no. 3, 2018, pp. 359–77. Taylor & Francis, https://doi.org/10.1080/2158379x.2018.1523315. Accessed 16 Aug. 2022.

---. "Shakespeare's Politics." *The Review of Politics*, vol. 78, no. 4, 2016, pp. 503–22. JSTOR, http://www.jstor.org/stable/24890014. Accessed 12 Jan. 2023.

Frey, Charles. "*The Tempest* and the New World." *Shakespeare Quarterly*, vol. 30, no. 1, 1979, pp. 29–41. JSTOR, https://doi.org/10.2307/2869659. Accessed 16 Aug. 2022.

Grady, Hugh. "Shakespeare's Links to Machiavelli and Montaigne: Constructing Intellectual Modernity in Early Modern Europe." *Comparative Literature*, vol. 52, no. 2, 2000, pp. 119–42. JSTOR, https://doi.org/10.2307/1771563. Accessed 16 Aug. 2022.

Laghi, Simona. "Utopias in *The Tempest*." *Pólemos*, vol. 11, no. 1, 2017, pp. 177–93. Proquest, https://doi.org/10.1515/pol-2017-0011. Accessed 16 Aug. 2022.

Machiavelli, Niccolò. *The Prince*. Translated by Harvey Mansfield, 2nd ed., University of Chicago Press, 1998.

Montaigne, Michel de. *The Complete Essays*. Penguin Books, 1993.

More, Thomas. *Utopia*. Edited by George M. Logan and Robert M. Adams, Cambridge University Press, 2002.

Oseman, Arlene. "The Machiavellian Prince in *The Tempest*." *Shakespeare in Southern Africa*, vol. 22, 2010, pp. 7-19. ProQuest, https://www.proquest.com/docview/1013751069. Accessed 16 Aug. 2022.

Samson, Alexander. "Introduction Locus Amoenus: Gardens and Horticulture in the Renaissance." *Renaissance Studies*, vol. 25, no. 1, 2011, pp. 1–23. JSTOR, https://doi.org/10.1111/j.1477-4658.2010.00714.x. Accessed 16 Aug. 2022.

Sanderlin, George. "The Meaning of Thomas More's 'Utopia.'" *College English*, vol. 12, no. 2, 1950, pp. 74–77. JSTOR, https://doi.org/10.2307/372227. Accessed 16 Aug. 2022.

Shakespeare, William. *The Tempest (Folger Shakespeare Library)*. Edited by Barbara A. Mowat and Paul Werstine, 1st ed., Simon and Schuster, 2004.

Stanivukovic, Goran. "*The Tempest* and the Discontents of Humanism." *Philological Quarterly*, vol. 85, no. 1, 2006, pp. 91-119. ProQuest, https://www.proquest.com/docview/211232597. Accessed 16 Aug. 2022.

Wells, Robin Headlam. *Shakespeare, Politics and the State*. Macmillan, 1986.

A Proposal for a Lipidomic Analysis of Cerebrospinal Fluid in Patients with Multiple System Atrophy

Siddharth Bhagwat

Author Background: *Siddharth Bhagwat grew up in the United States and currently attends Flower Mound High School in Flower Mound, Texas in the United States. His Pioneer research concentration was in the field of neuroscience and titled "The Brain Under Attack."*

1. Introduction

Multiple system atrophy (MSA) is an adult-onset orphan neurodegenerative disease marked by severe and rapidly progressing dysfunction of the nervous system. Pathogenesis of MSA is characterized by the misfolding and aggregation of the α-synuclein protein, both within neurons and, more prominently, within oligodendrocytes, where these aggregates form glial cytoplasmic inclusions (GCIs) (Fanciulli et al., 2019; Papp et al., 1989).

As disease progression is extremely rapid, life expectancy is short, typically ranging from 6 to 9 years after the onset of symptoms. The disease is relatively rare, with .6 cases per 100,000 people in the general populace, but increases in prevalence with age to a rate of 3 per 100,000 people over 50 (Krismer & Wenning, 2017). Presently, understanding of MSA is poor, and further study is needed to elucidate its pathophysiology and develop effective treatments and biomarkers.

1.1. Pathology of Multiple System Atrophy

Characterized by the build-up and aggregation of α-synuclein (α-syn) within the patient's cells, MSA is considered a synucleinopathy, as are Parkinson's disease (PD) and dementia with Lewy bodies. Notably, unlike in other synucleinopathies, these aggregates are observed in both the patient's neurons and their oligodendrocytes as GCIs (Fanciulli et al., 2019). This also raises questions about the pathophysiology of multiple system atrophy, as even in patients with the disease, oligodendrocytes do not express α-syn. Another differentiator between MSA and other synucleinopathies is the variation in the conformational strains of α-syn. It has been demonstrated that α-syn in MSA has a different structure and that aggregates of α-syn in MSA are more toxic than those of α-syn in PD (Shahnawaz et al., 2020). This difference in conformation could perhaps help explain some of the varying pathology observed in MSA versus PD.

Despite these hallmarks of MSA, the disease is rather varied in its clinical presentation, potentially making diagnosis difficult. A certain diagnosis can only be established post-mortem with the observation of GCIs (Gilman et al., 2008) . There are two primary varieties of multiple system atrophy: multiple system atrophy with predominant parkinsonism (MSA-P) and multiple system atrophy with predominant cerebellar ataxia (MSA-C). However, the disease can present clinically in a variety of other ways, including autonomic nervous system dysfunction (MSA-A) and upper motor neuron disease (UMN) (Brettschneider et al., 2018). Furthermore, the disease can be categorized by its morphology into two types, olivopontocerebellar dysfunction (OPCA), which typically corresponds to the MSA-C phenotype, or striatonigral degeneration (SND), which is similar in nature to MSA-P. Although initial clinical presentation often differs between variants of MSA, as the disease progresses, the phenotypic variation observed in patients with different categorizations of the disease wanes (Gilman et al., 2008).

Although MSA-P and MSA-C are now the most widely used categorizations of MSA (Gilman et al., 2008), the variations in how the disease's phenotypes were previously classified remain prevalent in biorepositories, including in one utilized in our proposal. Because of this, the lipidomic analysis of each cohort may not be entirely applicable to the others which may have been classified differently.

1.2. Lipids, α-synuclein, and Multiple System Atrophy

Lipids are crucial to the function of the brain, playing a variety of roles ranging from organization to energy storage to signal transduction (Taghibiglou et al., 2017). Lipid regulation is significantly different in patients with multiple system atrophy. Cao et al. (2014) found that patients with MSA had significantly reduced serum lipid levels. These changes also applied when considering cholesterol alone, with MSA patients having significantly lower high-density lipoprotein and total cholesterol than matched controls (Lee et al., 2009).

Giannakis et al. (2008) demonstrated that the strain of α-syn implicated in MSA pathology preferentially binds to lipid bilayers. Since this discovery, numerous studies have been undertaken in PD to look into the effects of α-syn on cerebrospinal fluid (CSF) lipidome composition, and a variety of significant differences have been found.

Despite the extensive study of the CSF lipidome in PD, no such studies have been undertaken in MSA. This leaves a void in our knowledge, much to the detriment of efforts to provide effective treatment. For instance, due to the lack of suitable biomarkers, patients may be misdiagnosed, causing great stress to them and their families. Such misdiagnoses may also slow clinical efforts, as patients may be included or excluded from relevant clinical trials. This proposal aims to outline a methodology by which an understanding of how MSA impacts the CSF lipidome could be obtained. By developing our knowledge of this critical clinical feature of MSA, this study, if undertaken, would help to address this pertinent problem affecting patients with the disease.

2. Proposed Methodology and Research Strategy

Our proposed study aims to compile a comprehensive lipidome of the CSF of patients with multiple system atrophy. The usage of CSF over blood in evaluating the lipidome provides several important advantages. Firstly, CSF composition has been shown to hold significant predictive power over neurodegenerative disease pathology. For example, in Alzheimer's disease (AD), CSF levels of the amyloid-β and phosphorylated tau proteins were able to aid in the effective prediction of AD pathology (Toledo et al., 2013). Additionally, the central nervous system is bathed by the cerebrospinal fluid and is cut off from the blood by the blood-brain barrier, inhibiting the free exchange of water-soluble substances between the CSF and blood (Engelhardt & Sorokin, 2009). This means that compared to blood, the CSF may present a more accurate picture of the metabolic and chemical processes occurring within the brain, aiding in better evaluations of disease. Finally, the CSF's composition is strictly regulated by the choroid plexus and is extremely stable, regardless of fluctuations in blood composition (Telano & Baker, 2022). Due to its comparative stability and consistency, CSF proves to be a valuable resource for developing biomarkers and elucidating some of the mechanisms behind MSA. It should be noted, however, that the use of CSF carries certain disadvantages as opposed to blood, namely, that the lumbar puncture required to extract CSF is more risky, invasive, and provides a comparatively minute quantity of CSF (Nociti et al., 2022). Nonetheless, because of the advantages provided by CSF, for the purposes of our proposal, we believe it would be of more value to utilize CSF over blood in compiling this lipidome.

2.1. Usage of Biobanks

Due to its rarity, one of the critical difficulties in studying MSA is finding a reliable and large patient cohort. In order to circumvent this issue and have well-documented and easily accessible data, we elected to develop a methodology involving the usage of biorepositories containing samples of cerebrospinal fluid from patients with multiple system atrophy.

We propose the usage of two major biorepositories of neurodegenerative disease cerebrospinal fluid to obtain samples of MSA CSF: the University of Pennsylvania Integrated Neurodegenerative Disease Biobank (INDD) and the Catalan MSA Registry (CMSAR). Both are large and well-maintained repositories with standardized protocols and storage procedures. Additionally, we propose the usage of a database from the Alzheimer's Disease Neuroimaging Initiative (ADNI) to provide matched controls for patients within the INDD as, unlike with the CSMAR cohort, our INDD cohort does not possess controls. All three repositories collected data with informed consent and permission of their respective regulatory boards.

2.2. The University of Pennsylvania Integrated Neurodegenerative Disease Biobank

We propose the usage of a sample within the Integrated Neurogenerative Disease Biobank (INDD) in this study. The INDD comprises a large and reliable repository

of biosamples from patients with neurodegenerative diseases. All patients included in the repository have a definitive neuropathological diagnosis of their respective conditions (Toledo et al., 2013). To ensure the accuracy of data entry, a variety of steps are taken. Randomly-selected data requires double entry, reducing the risk of incorrect numbers, and 10% of data is verified with the original source records every quarter.

Information regarding the type, date, and physical location of the samples are listed in the INDD (Toledo et al., 2013). Furthermore, each sample tube is labeled with the center at which the sample was collected and the processing date. Clinical characteristics of patients are also available within the database (Brettschneider et al., 2018). The information available within the INDD has been summarized in Figure 1.

Unfortunately, as there is no publicly available information, it is unknown whether data regarding disease progression is available within the INDD. If such data were to be available, we would analyze it in the manner outlined in the Statistical Methods section.

The INDD has standardized procedures for the procurement and processing of biosamples. Firstly, all biosamples are processed on the same day they are obtained (Toledo et al., 2013). Additionally, to reduce the probability of confounders originating in the patient's diet, all CSF samples are taken after at least 4 hours of fasting. CSF samples are immediately divided, sealed, stored on dry ice, and sent to their respective research divisions. The samples are stored long-term at -80℃ in specialized freezers with the sole purpose of storing biosamples. In order to verify minimal contamination by blood, we would utilize a Combur10-Test. We would reject samples with a reading of greater than 3+ (+++), as this corresponds to a blood concentration level of 50 RBC/µL, at which a sample is considered contaminated (Barkovits et al., 2020; Teunissen et al., 2009).

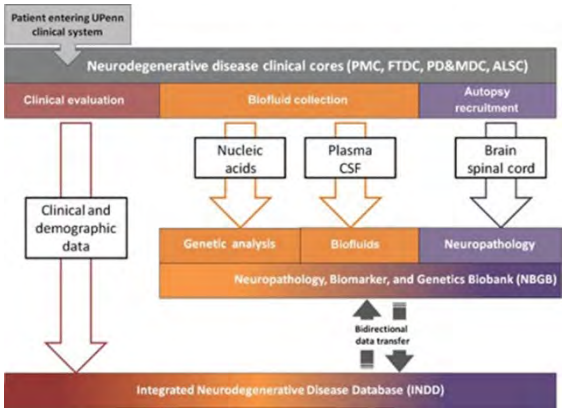


Figure 1. *Flowchart demonstrating data available in the INDD (Toledo et al., 2013)*

2.3. The Alzheimer's Disease Neuroimaging Initiative

Due to the lack of a known control population in the INDD, we elected to propose the use of normal (neurologically unimpaired) controls from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The particular site of the ADNI that we will draw controls from is the University of Pennsylvania's ADNI Biomarker Core laboratory.

The ADNI possesses a large number of biosamples from healthy controls, patients with mild cognitive impairment, and patients with Alzheimer's disease (Shaw et al., 2009). Much like with the CMSAR, patients are followed longitudinally and have biosamples, including CSF, extracted at 6, 12, 24, and 36 months (Weiner et al., 2014). Additionally, demographic and clinical data are available in the ADNI database (Shaw et al., 2009).

CSF samples were extracted using a lumbar puncture following a night of fasting (Shaw et al., 2009). The CSF was aliquoted and frozen on dry ice within one hour of collection before being transported overnight to the ADNI biomarker core laboratory, where it was frozen at -80°C and stored long-term. To evaluate blood contamination, we would once again utilize the procedure outlined in The University of Pennsylvania Integrated Neurodegenerative Disease Biobank section.

2.4. Catalan Multiple System Atrophy Registry

The third biorepository included in this proposal is the Catalan Multiple System Atrophy Registry (CMSAR). Like the INDD, the CMSAR has a relatively large and well-maintained store of biosamples from patients with MSA. Additionally, like the INDD, demographic and epidemiological data on patients is available in the database (Antonelli et al., 2016). Patients with MSA are also matched with neurologically unimpaired controls of the same age and sex (Pérez-Soriano et al., 2020).

Patients in this database were diagnosed according to the second consensus statement on the diagnosis of multiple system atrophy (Compta et al., 2019). Patients were then followed longitudinally, with biosamples being taken every 6 months (Antonelli et al., 2016). Additionally, information regarding disease progression in the form of Unified Multiple System Atrophy Rating Scale scores is available for all patients contained within the CMSAR (Compta et al., 2019). This provides the important advantage of understanding how CSF varies with the progression of the disease.

CSF samples in the cohort were obtained between 8 and 10 AM following a night of fasting (Compta et al., 2019). The CSF was then immediately centrifuged, aliquoted, and stored at a temperature of -80°C. To evaluate blood contamination, we would once again utilize the procedure outlined in The University of Pennsylvania Integrated Neurodegenerative Disease Biobank section.

2.5. Cohort Design and Statistics

We utilized discovery and validation cohorts in our proposal to verify our results.

Due to its more reliable status of diagnosis, we proposed the utilization of our sample from the INDD along with our control sample from the ADNI as our discovery cohort, and a sample from the CMSAR as our validation cohort.

Due to the small sample sizes, we found it infeasible to eliminate patients from the study altogether due to a potential lack of matching controls. Because of this, further research would be necessary to validate any findings.

Our first sample, which would be used as our discovery cohort, consists of 47 patients contained within the INDD, all with neuropathologically confirmed cases of MSA (Brettschneider et al., 2018). As 5 patients lacked age data, we would be unable to provide matching controls for all of the patients. Additionally, some patients were below the age of 50, making it challenging to pair them with healthy controls, given that the ADNI only consists of subjects aged 55 and older (Weiner et al., 2014). As a result, we would pair these patients with controls of the same sex and age 55. After matching, our discovery cohort would consist of 47 patients and 42 matched controls.

Our validation cohort would consist of 39 patients from the CMSAR with MSA (Compta et al., 2019). As all patients within the CMSAR are matched, using age and sex, with healthy controls, we would use the CMSAR to source 39 healthy controls (Antonelli et al., 2016). Due to the availability of matched controls, no special steps would be needed to develop the control group.

In our discovery cohort, 4 patients presented with MSA-A, 26 presented with MSA-P, 9 presented with MSA-C, 1 presented with UMN, and 6 patients lacked data regarding their initial clinical presentation (Brettschneider et al., 2018). Additionally, our INDD cohort was also categorized into pathological subtypes based on the damage observed in the brain, being of either the SND type, of which there were 24 patients, or the OPCA type, of which there were 10 patients. In our validation cohort of 39 patients, 20 presented with MSA-P and 19 presented with MSA-C (Compta et al., 2019). These data are represented graphically in Figure 2 and Tables 1 and 2. Additionally, relevant demographic data regarding the discovery and validation cohorts are available in Tables 1 and 2, respectively.

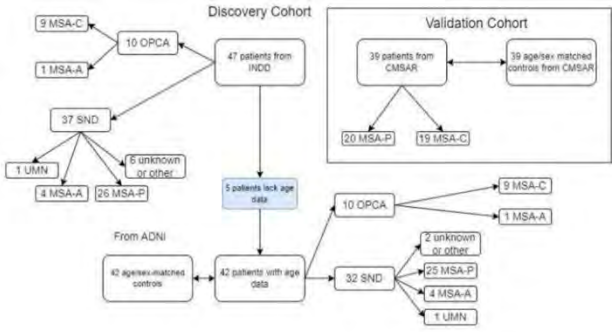


Figure 2. *Flowchart demonstrating categorization of patients in discovery and validation cohorts*

Table 1. *Demographic information of discovery cohort experimental group (Brettschneider et al., 2018)*[1]

	Total	SND	OPCA	MSA-P	MSA-C	MSA-A	UMN
Sample Size	47	37	10	26	9	5	1
Number of Women	16	15	1	N/A	N/A	N/A	N/A
Median Age at Onset (IQR)	56 (51 -63)	58 (49.75 - 63.25)	54.5 (53 - 60.25)	60 (54 - 64)	54 (53 - 62)	55 (49 - 56)	70 (N/A)
Median Disease Duration (IQR)	72 (48 - 108)	84 (48 -103)	57.5 (48 - 81)	84 (60 - 108)	66 (48 - 84)	48 (45 - 60)	48 (N/A)

Table 2. *Demographic information of validation cohort experimental group (Compta et al., 2019)*

	Total	MSA-P	MSA-C
Sample Size	39	19	20
Number of Women	17	8	9
Median Age at Onset (IQR)	57 (50 - 62)	57 (48 - 63)	56 (51 - 60)
Median Disease Duration (IQR)	62 (55 - 66)	64 (55 - 69)	61 (55 - 69)

2.6. Lipidomic Analysis Methods

To direct our CSF lipidome, we proposed to take guidance from Sachrione et al. (2021) and analyze many of the same major lipid classes that were previously studied in PD. We reason that this will better enable us to make meaningful comparisons between lipid levels observed in MSA and those in PD, which is also a synucleinopathy. However, we also would analyze the CSF concentration of various additional lipids that we thought might provide insightful results. The list of lipid classes that would be analyzed and their respective justifications for inclusion is provided in Table 3.

Lipidomics work would be outsourced to the Biomarkers Core Laboratory at Columbia University. The Biomarkers Core Laboratory would run a targeted lipidomics assay, which was selected due to its ability to analyze 34 lipid classes and 593 individual lipid species ("Targeted Lipidomics," 2022). Another key advantage of using the Biomarkers Core Laboratory and this particular assay would be the high sensitivity and accuracy provided. This advantage is obtained by the laboratory's usage of a liquid chromatography-mass spectrometry setup, which is extremely sensitive in its detection of lipids (Wang et al., 2019).

Lipids would be extracted from samples using a modified version of the Bligh-Dyer method ("Targeted Lipidomics", 2022). Firstly, CSF samples would be spiked with internal standards, which would allow us to account for variability in results caused by the processing and analysis of the samples (Mullaugh, 2020). Then, the spiked samples would be homogenized with a mixture of chloroform and methanol before being diluted with more chloroform and distilled water. This would separate the homogenate into a chloroform layer, consisting of the lipid components, and a methanol layer, consisting of the non-lipid components (Bligh & Dyer, 1959).

[1] Some patients were excluded due to a lack of data.

Table 3. *Table of major lipid classes included in the proposed lipidomic analysis*

Lipid Type		Reason for Inclusion
Glycerophospholipids	Phosphatidylcholine	A bilayer consisting of a type of phosphatidylcholine was found to catalyze the aggregation of α-syn into dimers. This occurs at a slower rate than for phosphatidylserine (Lv et al., 2019; Sachrione et al., 2021).
	Phosphatidylethanolamine	Disruption of an enzyme implicated in phosphatidylethanolamine synthesis resulted in cytoplasmic accumulation of α-syn in yeast and worm models of PD (Wang et al., 2014; Sachrione et al., 2021).
	Phosphatidylinositol	Phosphatidylinositol has been found to be decreased in yeast, rat, and human cortical neurons that are overexpressing α-syn (Sachrione et al., 2021).
	Cardiolipin	Ramakrishnan et al. (2003) found through spin labeling that cardiolipins interact with α-syn.
	Phosphatidic acid	Variation in the gene encoding diacyglycerol kinase theta, an enzyme involved in the production of phosphatidic acid from diaglycerols, have been associated with an increased susceptibility to PD (Xicoy et al., 2019).
	Phosphatidylglycerol	Phosphatidylglycerol has been found to interact with α-syn (Ramakrishnan et al., 2003).
	Phosphatidylserine	A bilayer consisting of a type of phosphatidylserine was found to catalyze the aggregation of α-syn into dimers. This occurred at a faster rate than the aforemention phosphatidylcholine bilayer (Lv et al., 2019; Sachrione et al., 2021). Phosphatidylserine also modulates interactions between SNARE-dependent vesicles and α-syn (Sachrione et al., 2021).
Sphingolipids	Sphingomyelin	Treatment of cells with exogenous sphingomyelin have been shown to increase levels of α-syn (Sachrione et al., 2021).
	Gangliosides	Exosomes containing gangliosides GM1 or GM3 were found to accelerate the process of α-syn aggregation (Sachrione et al., 2021).
	Cerebrosides	Mutations in glucocerebrosidase, an enzyme that degrades cerebrosides, are known risk factors for PD, another synucleinopathy (Zunke et al., 2018).
	Ceramides	Various inhibitors of ceramide synthesis have been found to greatly enhance the toxicity of α-syn in yeast models (Lee et al., 2011).
Saturated Fatty Acids	Stearic acid	Stearic acid interacts with α-syn (Sachrione et al., 2021).
	Myristic acid	One of the most common saturated fatty acids in the brain (Fecchio et al., 2018).
	Lignoceric acid	
	Lauric acid	
	Palmitic acid	A diet rich in palmitic acid was found to elevate α-syn expression in mice models of PD (Schommer et al., 2018).
Unsaturated Fatty Acids	Palmitoleic acid	Has been found to be decreased in PD CSF (Sachrione et al., 2021).
	α-linolenic acid	Has been found to promote formation of α-syn oligomers (Sachrione et al., 2021).
	Arachidonic acid	Together with docosahexaenoic acid, this lipid comrpises around 20% of the fatty acids in the brain (Rapoport, 2008).
	Oleic acid	Inhibition of stearoyl-CoA-desaturase, an enzyme involved in the production of oleic acid, was protective against PD in model organisms. Oleic acid was also found to promote α-syn aggregation (Fanning et al., 2019).
Omega-3 Unsaturated Fatty Acids	Eicosapentaenoic acid	Found to be reduced in lipid rafts during PD (Sachrione et al., 2021).
	Docosahexaenoic acid	Docosahexaenoic acid has been found to modulate α-syn oligomerization (Sachrione et al., 2021).
Sterols	Free Cholesterol	Serum cholesterol levels were found to be significantly reduced in patients with MSA (Lee et al., 2009).
	Cholesterol Esters	
Glycerolipids	Monoacylglycerols	Variation in expression of monoacylglycerol lipase, an enzyme that degrades monoacylglycerols, has been observed in PD patients and is thought to be related to mechanisms of neurodegeneration (Xicoy et al., 2019).
	Diacylglycerols	Variations in the gene encoding diacylglycerol kinase theta, an enzyme involved in the production of phosphatidic acid from diacylglycerols, have been associated with an increased susceptibility to PD. PD patients also have been shown to have increased levels of diacylglycerols in their frontal cortices (Xicoy et al., 2019).
	Triacylglycerols	Increased triacylglycerol levels have been correlated with a reduced risk of PD. Interestingly, triacylglycerol levels have been found to have been decreased in serum and plasma of male PD patients (Xicoy et al., 2019).

The extracted lipids would then be separated using high-performance liquid chromatography (HPLC) on an Agilent 1260 Infinity platform ("Targeted Lipidomics", 2022). To aid in separation, normal phase HPLC would be used for glycerophospholipids and sphingolipids, whereas for sterols, fatty acids, and glycerolipids, a reverse phase procedure would be used (Christie, n.d.; "Targeted Lipidomics", 2022). After separation, lipid levels would be quantified with multiple reaction monitoring mass spectrometry using an Agilent 6490 Triple Quadrupole mass spectrometer.

2.7. Statistical Methods

After collecting data on the CSF lipidome of each patient, we would compile averages of CSF lipid concentration for both the controls and experimental groups in each population. We would then run a variety of statistical tests of significance to evaluate any differences. The specific details on the tests utilized are discussed in greater detail below. Throughout, a cutoff of $p = .05$ would be utilized. With additional data, it could also be informative to take into consideration other factors, such as false discovery rates, before deciding the significance (Grabowski, 2016). These other factors could also provide grounds for further studies in this area.

In brief, to compare the MSA population to the control population, we would run significance tests evaluating the differences between the mean levels of lipids in our MSA patients and those in our controls. We would analyze these differences for both each individual lipid subclass (e.g., phosphatidylcholine) and each broader lipid group (e.g., glycerophospholipids). All tests utilized are robust to violations of normality, allowing the same tests to be used in the case that the data do not resemble a normal distribution (Blanca et al., 2017; Olejnik & Algina, 1984; Snijders, 2011).

To compare our discovery cohort to our control population, we would utilize an analysis of covariance (ANCOVA). Using this, we would test for significant differences in lipid concentrations in each class and subclass between the MSA population and the control population. This test enables us to evaluate these differences while accounting for the effects of age as a potential confounder. To do this, we would first have to exclude the 5 patients from our INDD cohort who lacked age-related data. Proceeding with the 42 remaining patients, we would utilize the ANCOVA to determine whether the variation between the MSA and control lipidome was significant, considering the influence of age. We would not run this test for our other populations, however, as there are no important confounders that would be addressed through the usage of an ANCOVA. Then, in an effort to confirm any results found in our discovery cohort, we would utilize Student's t-test to compare lipid concentrations for each class and subclass between our validation cohort's experimental group and its respective control.

Furthermore, as our validation cohort is known to have data regarding disease progression, we would also categorize samples by the patient's corresponding Unified Multiple System Atrophy Rating Scale and utilize a one-way analysis of variance (ANOVA) to evaluate differences between these categories. This would enable us to develop insight into alterations in the lipidome that occur with the disease's progression. If data on disease progression

were available in our discovery cohort, we would analyze this in the same manner.

In both our validation cohorts and our discovery cohorts, we would use t-tests to look for significant differences between our MSA-P and MSA-C samples overall.

In our validation cohort, we would once again categorize samples by Unified Multiple System Atrophy Rating Scale to better understand how differences in the lipidome between MSA-P and MSA-C are affected by the stage of illness. To perform this analysis, we would utilize a two-way ANOVA. As was the case with our previous analysis of disease progression, we would perform a corresponding analysis in our discovery cohort if such data was available. We would also analyze disease progression in MSA-A and UMN, although this would not have a counterpart in the validation cohort.

Finally, in our discovery cohort, we would utilize a one-way ANOVA to compare the lipidome in MSA-P, MSA-C, MSA-A, and UMN. It should be noted that as is the case in a potential analysis of MSA progression in our discovery cohort, any significant differences found between MSA-A or UMN and the other phenotypic variants in the INDD could not be confirmed with the validation cohort owing to the different categorizations. A t-test comparing the SND and OPCA variants of MSA would also be run; however, as is the case of the aforementioned ANOVA, this test would not have a counterpart in the validation cohort.

3. Expected Outcomes

3.1. Objectives

By conducting a study in line with the previously outlined protocols, we would expect to develop a better understanding of how multiple system atrophy affects the cerebrospinal fluid lipidome, a vital component of our understanding of the disease. This could potentially have crucial implications for the discovery of biomarkers and the elucidation of MSA's pathogenesis.

For example, as the CSF lipidome in PD is relatively well-known, a potential difference in lipid concentrations could perhaps be used as a diagnostic tool to increase the certainty of MSA diagnosis prior to death. Additionally, such a difference could also provide clues as to how the pathology of PD and MSA differ.

3.2. Hypothesized Results

As per the Second Consensus Statement on the Diagnosis of MSA, the variation in the clinical presentation of the variants of MSA wanes as the disease progresses, converging to a common phenotype (Gilman et al., 2008). Similarly, we hypothesize that if there are lipidomic differences between the different variants of MSA, this variation should reduce as progression continues.

A tenet of MSA pathology is the observance of GCIs and subsequent oligodendrocyte dysfunction, leading to dysfunction in myelin metabolism (Bleasel et al., 2014). This defective regulation of myelin metabolism could result in below-normal levels of the lipid components of the Myelin sheath, such as

glycolipids, phospholipids, and cholesterols (Poitelon, 2020).

Finally, since both MSA and PD are synucleinopathies, we would expect the lipidomic profile of both diseases to be quite similar. Fernández-Irigoyen et al. (2021) demonstrated that levels of glycerolipids, primary fatty amides, cholesterol esters, steroids, saturated fatty acids, phosphatidylcholines, phosphatidylethanolamines, and N-acylethanolamines were significantly increased in patients with PD. Similar changes in the CSF lipidome could also be observed in MSA. It should be noted that some of these observed changes in PD contradict those that we would expect in MSA, such as a reduction in cholesterols and glycolipids.

3.3. Potential Concerns

There are several limitations concerning the methodology of this study. Firstly, the patients included in the validation cohort, unlike those in the discovery cohort, lack a definite neuropathological diagnosis of MSA (Compta et al., 2019). This carries with it the risk of an improper diagnosis being included in the cohort, potentially clouding results. Unfortunately, due to the lack of biorepositories, we were unable to circumvent this risk regarding the use of non-neuropathologically confirmed patients.

Nomenclature also provided some difficulty in analyzing the results. As previously mentioned, MSA can be categorized phenotypically into MSA-P, MSA-C, and others, or morphologically into MSA with either SND or OPCA. As this lack of standardized terminology had the potential to influence our results, we elected to analyze all of the various clinical presentations of MSA in our two cohorts. To combat this, future studies could draw on databases containing only the categorizations of MSA-P and MSA-C, as is standard with modern diagnostic protocols for MSA (Gilman et al., 2008). Additionally, if possible, with the collected information, future studies could attempt to recategorize patients into either MSA-P or MSA-C using the current consensus for the diagnosis of MSA.

Another concern that should be noted is the lack of standardized procedures in the collection and processing of samples. As the biorepositories used were either not affiliated or only loosely affiliated, each used its own unique procedures in collecting biosamples. To stress these differences, the sample collection and processing methods utilized at each biobank were listed in the ‘Sample Collection’ subsection of the ‘Proposed Methodology and Research Strategy’ section.

Due to the lack of available data, there are also a variety of demographic constraints associated with this proposal. As previously discussed, the age difference between the control group contained in the ADNI and the experimental group within the INDD would have the potential to confound our results. Although we would attempt to compensate for these effects by using an ANCOVA to test for significant differences, it would have been better to use matched controls in the first place.

Perhaps more concerning, though, is the shortage of women in our INDD cohort with the OPCA phenotype and, as a product, MSA-C. This shortage underscores the need for our validation cohort and may warrant further study, considering the high potential for confounders.

Finally, another crucial limitation of this proposal is the potential lack of

insight gained into how the disease lipidome changes with time. Although our validation cohort does have this data, our discovery cohort may not, and any results found may not have the benefit of confirmation with another cohort within the study. This limits much of our proposal to the overall changes in the lipidome that occur with MSA and underscores the need for better developed resources for the disease's study.

Nonetheless, given the mainly exploratory nature of this study, these concerns are outweighed by the benefits conveyed by the use of a relatively large sample.

3.4. Conclusion

In this proposal, we outlined the methodology and necessity of a comprehensive lipidomic study. We discussed in detail the cohorts that would be utilized and our reasoning behind their usage. We also explained the various statistical strategies we would use to analyze our data. Finally, we discussed the potential concerns facing this proposal and the methods by which we would try to account for them.

The previously outlined concerns help highlight the necessity for more standardized procedures for future studies on MSA. Furthermore, even if this study were to be carried out to completion, the need would remain for larger and more precise studies to help confirm results and apply this newfound knowledge to the search for biomarkers and treatments for MSA.

References

Antonelli, F., Muñoz, E., Pagonabarraga, J., Hernández-Vara, J., Bayes, A., de Fabregues, O., Valldeoriola, F., Tolosa, E., Compta, Y., Ezquerra, M., Fernandez, R., Calopa, M., Jauma, S., Pujol, M., Puente, V., Cámara, A., Planellas, L., Martí, M.J. (2016). The Catalan multiple system atrophy-registry (CMSAR) [abstract]. Mov Disord.; 31 (suppl 2). https://www.mdsabstracts.org/abstract/the-catalan-multiple-system-atrophy-registry-cmsar/

Barkovits, K., Kruse, N., Linden, A., Tönges, L., Pfeiffer, K., Mollenhauer, B., & Marcus, K. (2020). Blood contamination in CSF and its impact on quantitative analysis of alpha-Synuclein. *Cells*, 9(2), 370. https://doi.org/10.3390/cells9020370

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? Psicothema, 29(4), 552–557. https://doi.org/10.7334/psicothema2016.383

Bleasel, J. M., Wong, J. H., Halliday, G. M., & Kim, W. S. (2014). Lipid dysfunction and pathogenesis of multiple system atrophy. Acta neuropathologica communications, 2, 15. https://doi.org/10.1186/2051-5960-2-15

Bligh, E. G., & Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Canadian journal of biochemistry and physiology*, 37(8), 911–917. https://doi.org/10.1139/o59-099

Brettschneider, J., Suh, E., Robinson, J. L., Fang, L., Lee, E. B., Irwin, D. J., Grossman, M., Van Deerlin, V. M., Lee, V. M., & Trojanowski, J. Q. (2018). Converging patterns of α-synuclein pathology in multiple system atrophy. *Journal of neuropathology and experimental neurology*, 77 (11), 1005–1016. https://doi.org/10.1093/jnen/nly080

Cao, B., Guo, X., Chen, K., Song, W., Huang, R., Wei, Q. Q., Zhao, B., & Shang, H. F. (2014). Serum lipid levels are associated with the prevalence but not with the disease progression of multiple system atrophy in a Chinese population. *Neurological research*, 36(2), 150–156. https://doi.org/ 10.1179/1743132813Y.0000000277

Christie, W. W. (n.d.). Fatty Acid Analysis by HPLC. AOCS Lipid Library. Retrieved August 28, 2022, from https://lipidlibrary.aocs.org/lipid-analysis/selected-topics-in-the-analysis-of-lipids/fatty-acid-analysis-by-hplc

Compta, Y., Dias, S. P., Giraldo, D. M., Pérez-Soriano, A., Muñoz, E., Saura, J., Fernández, M., Bravo, P., Cámara, A., Pulido-Salgado, M., Painous, C., Ríos, J., & Martí, M. J. (2019, June 3). Cerebrospinal fluid cytokines in multiple system atrophy: A cross-sectional Catalan MSA Registry study. *Parkinsonism & Related Disorders*. Retrieved August 9, 2022, from https://www.sciencedirect.com/science/article/pii/S1353802019302615?via%3Dihub

Engelhardt, B., & Sorokin, L. (2009). The blood-brain and the blood-cerebrospinal fluid barriers: function and dysfunction. *Seminars in immunopathology*, 31(4), 497–511. https://doi.org/10.1007/s00281-009-0177-0

Fanciulli, A., Stankovic, I., Krismer, F., Seppi, K., Levin, J., & Wenning, G. K. (2019, November 21). Multiple system atrophy. *International Review of Neurobiology*. Retrieved August 3, 2022, from https://www. Science direct.com/science/article/pii/S007477421930090X?via%3Dihub

Fanning, S., Haque, A., Imberdis, T., Baru, V., Barrasa, M. I., Nuber, S., Termine, D., Ramalingam, N., Ho, G., Noble, T., Sandoe, J., Lou, Y., Landgraf, D., Freyzon, Y., Newby, G., Soldner, F., Terry-Kantor, E., Kim, T. E., Hofbauer, H. F., Becuwe, M., ... Selkoe, D. (2019). Lipidomic analysis of α-synuclein neurotoxicity identifies stearoyl CoA desaturase as a aarget for Parkinson treatment. *Molecular cell*, 73(5), 1001–1014.e8. https://doi.org/10.1016/j.molcel.2018.11.028

Fecchio, C., Palazzi, L., & de Laureto, P. P. (2018). α-Synuclein and polyunsaturated fatty acids: molecular basis of the interaction and implication in neurodegeneration. *Molecules* (Basel, Switzerland), 23(7), 1531. https://doi.org/10.3390/ molecules23071531

Fernández-Irigoyen, J., Cartas-Cejudo, P., Iruarrizaga-Lejarreta, M., & Santamaría, E. (2021). Alteration in the cerebrospinal fluid lipidome in Parkinson's disease: A post-mortem pilot study. *Biomedicines*, 9(5), 491. https://doi.org/10.3390/biomedicines9050491

Giannakis, E., Pacífico, J., Smith, D. P., Hung, L. W., Masters, C. L., Cappai, R., Wade, J. D., & Barnham, K. J. (2008). Dimeric structures of alpha-synuclein bind preferentially to lipid membranes. *Biochimica et biophysica acta*, 1778(4), 1112–1119. https://doi.org/10.1016/j.bbamem.2008.01.012

Gilman, S., Wenning, G. K., Low, P. A., Brooks, D. J., Mathias, C. J., Trojanowski, J. Q., Wood, N. W., Colosimo, C., Dürr, A., Fowler, C. J., Kaufmann, H., Klockgether, T., Lees, A., Poewe, W., Quinn, N., Revesz, T., Robertson, D., Sandroni, P., Seppi, K., & Vidailhet, M. (2008, August 26). Second consensus statement on the diagnosis of multiple system atrophy. *Neurology*. Retrieved August 3, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2676993/

Grabowski B. (2016). "P < 0.05" might not mean what you think: American statistical association clarifies P values. *Journal of the national cancer institute*, 108(8), djw194. https://doi.org/10.1093/jnci/djw194

Irving Institute for Clinical and Translational Research (2022, January 18). Targeted lipidomics. Columbia University Irving Medical Center. Retrieved August 9, 2022, from https://www.irvinginstitute.columbia.edu/services/targeted-lipidomics

Krismer, F., & Wenning, G. K. (2017, March 17). Multiple system atrophy: Insights into a rare and debilitating movement disorder. *Nature news*. Retrieved August 3, 2022, from https://www.nature.com/articles/nrneurol.2017.26

Lee, P. H., Lim, T. S., Shin, H. W., Yong, S. W., Nam, H. S., & Sohn, Y. H. (2009). Serum cholesterol levels and the risk of multiple system atrophy: a case-control study. *Movement disorders: official journal of the Movement Disorder Society*, 24(5), 752–758. https://doi.org/10.1002/mds.22459

Lee, Y. J., Wang, S., Slone, S. R., Yacoubian, T. A., & Witt, S. N. (2011). Defects in very long chain fatty acid synthesis enhance alpha-synuclein toxicity in a yeast model of Parkinson's disease. PloS one, 6(1), e15946. https://doi.org/10.1371/journal.pone.0015946

Lv, Z., Hashemi, M., Banerjee, S., Zagorski, K., Rochet, J. C., & Lyubchenko, Y. L. (2019). Assembly of α-synuclein aggregates on phospholipid bilayers. Biochimica et biophysica acta. *Proteins and proteomics*, 1867(9), 802–812. https://doi.org/10.1016/j.bbapap.2019.06.006

Mullaugh, K. (2020, October 20). Internal standards and Lod. Chemistry LibreTexts. Retrieved August 28, 2022, from https://chem.libretexts.org/Bookshelves/Organic_Chemistry/Organic_Chemistry_Lab_Techniques_(Nichols)/04%3A_Extraction/4.04%3A_Which_Layer_is_Which

Nociti, V., Romozzi, M., & Mirabella, M. (2022). Update on multiple sclerosis molecular biomarkers to monitor treatment effects. *Journal of personalized medicine*, 12(4), 549. https://doi.org/10.3390/jpm12040549

Olejnik, S. F., & Algina, J. (1984). Parametric Ancova and the rank transform Ancova when the data are conditionally non-normal and heteroscedastic. *Journal of Educational Statistics*, 9(2), 129–149. https://doi.org/10.3102/10769986009002129

Papp, M. I., Kahn, J. E., & Lantos, P. L. (1989). Glial cytoplasmic inclusions in the CNS of patients with multiple system atrophy (striatonigral degeneration, olivopontocerebellar atrophy and Shy-Drager syndrome). *Journal of the neurological sciences*, 94(1-3), 79–100. https://doi.org/10.1016/0022-510x(89)90219-0

Pérez-Soriano, A., Arnal Segura, M., Botta-Orfila, T., Giraldo, D., Fernández, M., Compta, Y., Fernández-Santiago, R., Ezquerra, M., Tartaglia, G. G., Martí, M. J., & Catalan MSA Registry (CMSAR) (2020). Transcriptomic differences in MSA clinical variants. *Scientific reports*, 10(1), 10310. https://doi.org/10.1038/s41598-020-66221-4

Poitelon, Y., Kopec, A. M., & Belin, S. (2020). Myelin fat facts: An overview of lipids and fatty acid metabolism. *Cells*, 9(4), 812. https://doi.org/10.3390/cells9040812

Rapoport S. I. (2008). Arachidonic acid and the brain. *The Journal of nutrition*, 138(12), 2515– 2520. https://doi.org/10.1093/jn/138.12.2515

Ramakrishnan, M., Jensen, P. H., & Marsh, D. (2003). Alpha-synuclein association with phosphatidylglycerol probed by lipid spin labels. *Biochemistry*, 42(44), 12919–12926. https://doi.org/10.1021/bi035048e

Sarchione, A., Marchand, A., Taymans, J. M., & Chartier-Harlin, M. C. (2021). Alpha-synuclein and lipids: The elephant in the room? *Cells*, 10(9), 2452. https://doi.org/10.3390/cells10092452

Schommer, J., Marwarha, G., Nagamoto-Combs, K., & Ghribi, O. (2018). Palmitic acid-enriched diet increases α-synuclein and tyrosine hydroxylase expression levels in the mouse brain. *Frontiers in neuroscience*, 12, 552. https://doi.org/10.3389/fnins.2018.00552

Shahnawaz, M., Mukherjee, A., Pritzkow, S., Mendez, N., Rabadia, P., Liu, X., Hu, B., Schmeichel, A., Singer, W., Wu, G., Tsai, A. L., Shirani, H., Nilsson, K., Low, P. A., & Soto, C. (2020). Discriminating α-synuclein strains in Parkinson's disease and multiple system atrophy. *Nature*, 578(7794), 273–277. https://doi.org/10.1038/s41586-020-1984-7

Shaw, L. M., Vanderstichele, H., Knapik-Czajka, M., Clark, C. M., Aisen, P. S., Petersen, R. C., Blennow, K., Soares, H., Simon, A., Lewczuk, P., Dean, R., Siemers, E., Potter, W., Lee, V. M., Trojanowski, J. Q., & Alzheimer's Disease Neuroimaging Initiative (2009). Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Annals of neurology*, 65(4), 403–413. https://doi.org/10.1002/ana.21610

Snijders, T. A. B. (2011, November 13). Statistical methods: Robustness - University of Oxford. University of Oxford - Department of Statistics. Retrieved August 25, 2022, from https://www.stats.ox.ac.uk/~snijders/SM_robustness.pdf

Taghibiglou, C., Khalaj, S., & Adejare, A. (2017). Chapter 9 - cholesterol and fat metabolism in alzheimer's disease. In Drug discovery approaches for the treatment of neurodegenerative disorders. essay, Academic Press.

Telano, L. L., & Baker, S. (2022, July 4). Physiology, cerebral spinal fluid - statpearls - NCBI bookshelf. National Library of Medicine. https://www.ncbi.nlm.nih.gov/books/NBK519007/

Teunissen, C. E., Petzold, A., Bennett, J. L., Berven, F. S., Brundin, L., Comabella, M., Franciotta, D., Frederiksen, J. L., Fleming, J. O., Furlan, R., Hintzen, MD, R. Q., Hughes, S. G., Johnson, M. H., Krasulova, E., Kuhle, J., Magnone, M. C., Rajda, C., Rejdak, K., Schmidt, H. K., ... Deisenhammer, F. (2009). A consensus protocol for the standardization of Cerebrospinal Fluid Collection and Biobanking. *Neurology*, 73(22), 1914–1922. https://doi.org/10.1212/wnl.0b013e3181c47cc2

Toledo, J. B., Van Deerlin, V. M., Lee, E. B., Suh, E., Baek, Y., Robinson, J. L., Xie, S. X., McBride, J., Wood, E. M., Schuck, T., Irwin, D. J., Gross, R. G., Hurtig, H., McCluskey, L., Elman, L., Karlawish, J., Schellenberg, G., Chen-Plotkin, A., Wolk, D., Grossman, M., ... Trojanowski, J. Q. (2014). A platform for discovery: The University of Pennsylvania Integrated Neurodegenerative Disease Biobank. *Alzheimer's & dementia: the journal of the Alzheimer's Association*, 10(4), 477–484.e1. https://doi.org/10.1016/j.jalz.2013.06.003

Wang, J., Wang, C., & Han, X. (2019). Tutorial on lipidomics. Analytica chimica acta, 1061, 28–41. https://doi.org/10.1016/j.aca.2019.01.043

Wang, S., Zhang, S., Liou, L. C., Ren, Q., Zhang, Z., Caldwell, G. A., Caldwell, K. A., & Witt, S. N. (2014). Phosphatidylethanolamine deficiency disrupts α-synuclein homeostasis in yeast and worm models of Parkinson disease. Proceedings of the National Academy of Sciences of the United States of America, 111(38), E3976–E3985. https://doi.org/10.1073/pnas. 1411694111

Weiner, M. W., Petersen, R., & Aisen, P. (2014, September 16). ADNI: Alzheimer's disease neuroimaging initiative. ADNI: Alzheimer's disease neuroimaging initiative. Retrieved August 9, 2022, from https://clinicaltrials.gov/ct2/show/NCT00106899

Xicoy, H., Wieringa, B., & Martens, G. (2019). The role of lipids in Parkinson's disease. Cells, 8(1), 27. https://doi.org/10.3390/cells8010027

Zunke, F., Moise, A. C., Belur, N. R., Gelyana, E., Stojkovska, I., Dzaferbegovic, H., Toker, N. J., Jeon, S., Fredriksen, K., & Mazzulli, J. R. (2018). Reversible conformational conversion of α-synuclein into toxic assemblies by glucosylceramide. Neuron, 97(1), 92–107.e10. https://doi.org/10.1016/j.neuron.2017.12.012

Chained to Knowledge? An Examination of Descartes' View on Free Will in the *Meditations*

Jiatong Liu

***Author background**: Jiatong Liu grew up in China and currently attends Keystone Academy in Beijing, China. Her Pioneer research concentration was in the field of philosophy and titled "Descartes' Meditations."*

1. Introduction

Descartes, the father of modern philosophy, marries theodicy and epistemology when he sets out to prove the existence of God by questioning all his preconditioned knowledge in his *Meditations on First Philosophy*. Observing his occasional errors in judgment, from sensory misperceptions to deceivingly realistic dreams, Descartes resolves to withhold assent for all preexisting beliefs and gradually builds his way up to a knowledge system with complete certainty. After he concludes the existence of an omnipotent God in the Third Meditation, one conceptual challenge remains. If there is a perfectly omnipotent God, why does He allow us to make errors in our judgment? Descartes attempts to answer this question in the Fourth Meditation by accounting for these errors with the existence of the human free will. God has given humans a perfectly functioning intellect and free will, but humans make errors when they extend their free will to matters that they do not understand. In the Fifth Meditation, this line of reasoning is developed further by the notion of a "clear and distinct" understanding, which sets apart the truth from error and ensures certainty in judgment. As "I am incapable of error in those cases where my understanding is transparently clear" (CSM II 70), Descartes constructs an apparent conflict between free will and "clear and distinct" knowledge. When we view something clearly and distinctly, do we lose control over our judgment? Consequently, does having knowledge undermine free will? If not, Descartes' whole foundation for theodicy is destabilized. But if so, Descartes challenges us to rethink the merits of our pursuit of knowledge. If free will and knowledge are incompatible, we may be forced to choose between the two virtues. Therefore, this paper aims to resolve this important tension between free will and certainty in knowledge in the *Meditations*.

In the first section, an overview of Descartes' understanding of freedom in the *Meditations* will be established, with reference to his *Principles* and Jesuit Letters. There are two main conflicting views of free will: freedom of indifference and freedom of spontaneity. After introducing these two lines of free will in Descartes' works, the paper will examine their relationship with the possession of "certain" knowledge: for example, the supposed threat of knowledge upon freedom of indifference, but not on freedom of spontaneity, in the *Meditations*. At the same time, the section will establish Descartes' potential response to this apparent tension between knowledge and free will. The simple objection in the *Meditations* and the *Principles* is dismissing

freedom of indifference, which is undermined by the possession of certain knowledge. The argument here still entails that certain knowledge presented by God predetermines judgment. A more involved objection is raised in the Jesuit Letters, as Descartes promotes the freedom of indifference to argue humans technically have choices even in the face of clear and distinct knowledge.

The objective of the second section is to examine Descartes' different lines of freedoms under compatibilist frameworks to contextualize these views in relation to judgment's divine preordination. As Descartes outwardly endorses the omnipotent power of God, a thought experiment will be introduced to challenge Descartes' compromised conception of libertarian free will in his *Meditations* and *Principles*. Deciding that the Jesuit Letters are more committed to true freedom, the paper will reexamine this freedom's compatibility with judgment's divine preordination.

The third section of the essay will discuss the wide-ranging implications of the resolved compatibilism between certain knowledge and free will in Descartes. Elaborating upon the valid part of Descartes' view on free will and judgments, the paper aims to pursue further inquiries on the question of divine freedom and the formation of false consciousness.

2. The Theodicy of Freedom

Under the epistemological context, Descartes' conception of human free will must be contextualized with its role in formulating judgments. From the *Meditations* to the *Principles*, Descartes repeatedly attributes the action of judging to two facilitating modes: the intellect and the will. The intellect allows for the perception and understanding of propositions – *that* such-and-such is the case – which could not be false unless we integrate judgment into a proposition by affirming or denying it. The number of propositions we can perceive clearly and distinctly is limited. The will, on the other hand, is the voluntary power within each individual to affirm or deny the propositions she perceives by the intellect. Our free will can be extended to all objects of propositions and thus can be deemed infinite. Both are given to us by God, our creator. In the *Meditation* as well as the *Principles*, Descartes maintains that humans err when we extend our immeasurable free will "to matters which [we] do not understand" by our limited intellect (CSM II 40).

This explanation of why we make mistakes builds a solid foundation for Descartes' theodicy. In the *Meditations*, with a distinction between two different notions of imperfection, Descartes showed that we cannot attribute to God our errors in judgment. God has given us immeasurable free will created in His own image, the most perfect of its kind. The constraint of our limited intellect prevents us from perceiving all ideas clearly and distinctly, and lacking features compared to the more perfect kind, our intellect is only imperfect in a negative sense (Newman 559). As positive imperfections are associated with defect and malfunctions, we evidently do not have any positive imperfections. So long as we only lack features by negative imperfection, there is no reason for us to blame God, as "I cannot produce any reason to prove that God ought to have given me a greater faculty of knowledge than he did" (CSM II 39). Descartes ultimately resolves the problem of evil in theodicy, as he demonstrated God and evil are compatible with a free will defense, explaining we have the freedom to avoid evil when we withhold our assent from unclear propositions.

Although in Descartes' mature philosophical works he consistently maintains that God is by nature all-good and that humans are naturally free, his conception of

freedom differs widely from text to text. To resolve the apparent conflict between free will and certain knowledge, it is necessary to first trace these diverging lines of freedom in making judgments.

2.1 Freedom in the *Mediations*

In his *Meditations*, Descartes presents two types of freedom: freedom of indifference and freedom of spontaneity. Before specifically writing about indifference and spontaneity, the Fourth Meditation first advances a claim subtly introducing these two concepts:

> The will simply consists in our ability to do or not do something (that is, to affirm or deny, to pursue or avoid); or rather, it consists simply in the fact that when the intellect puts something forward for affirmation or denial or for pursuit or avoidance, our inclinations are such that we do not feel we are determined by any external force. (CSM II 40)

Previous scholars have read this definition as two clauses for freedom (Kaufman 391). The first, emphasizing our "ability to do or not do something," implicates freedom of indifference, the underlying freedom to do otherwise in each choice made. The second suggests freedom of spontaneity, as the individual is free to choose, or incline toward, what she wants. It seems like both types of freedoms are prerequisites for true free will.

However, when we make judgments, these two types of freedom come into conflict with each other. Where we face clear and distinct perception by the intellect, we cannot but assent to it, losing the two-way power of doing otherwise. While our freedom of indifference is diminished, our freedom of spontaneity seems to increase greatly by inverse proportionality, as this lack of indifference adds to our one-way inclination. Where we face obscure and confused perceptions, "the will is indifferent" and errs as it "easily turns aside from what is true and good" in such cases (CSM II 41). Therefore, when freedom of indifference is present in judgments, one loses the freedom of spontaneity to incline for "what is true and good." This conflict is problematic as it suggests full freedom that meets the two requirements can never be achieved.

Descartes responds to this conflict by dismissing freedom of indifference as unimportant, or in his words, "the lowest grade of freedom" (CSM II 40). According to Descartes, "to be free, there is no need for me to be inclined both ways; on the contrary, the more I incline in one direction... the freer is my choice" (CSM II 40). On the surface puzzling, this claim is in fact grounded in Descartes' previous arguments. Indifference itself reflects the negative imperfection in us finite beings. If one's intellect knows everything clearly and distinctly, there would be no need to remain indifferent and one can transcend into higher grades of freedoms. As long as freedom of spontaneity is fostered by the possession of certain knowledge, the pursuit of certainty remains worthwhile. Therefore, in the *Meditations*, Descartes simply reconciles the two virtues of certain knowledge and free will by dismissing the indifferent part of the will as trifling and unworthy.

2.2 Freedom in the *Principles*

The *Principles* differs from the *Meditations*, as it treats the freedom of indifference

not as a defect but as a celebrated characteristic of humanity. In the text, Descartes perceives the freedom of indifference to withhold consent as what distinguishes human beings from automata: whereas we act freely, automatons act necessarily. When we come to the truth "voluntarily", as Descartes points out, is "much more to our credit than would be the case if we could not do otherwise" (CSM I 205). Although this quote seemingly suggests that freedom of indifference can survive even under clear and distinct perceptions, indifference is only hypothetical. Although this passage does not treat the freedom of indifference as a defect, it still insists on the very next page that all actions are preordained by God.

The text still largely agrees with the *Meditations* in the dominance of spontaneity whenever individuals are met with clear and distinct perceptions by the intellect. As our minds by nature "spontaneously give our assent" to any clear and distinct perception and "are quite unable to doubt its truth" (CSM I 207), these lines are reiterating the inevitability of assenting to clear and distinct perceptions in the *Meditations*.

Descartes reconciles free will and certain knowledge by the presence of moral responsibility. Although we spontaneously give our assent to "what is true and good," in hypothesis, we can withhold judgment as free agents. Although predetermined, the choice is still free in the sense that it results from human free will or at least the illusion of being able to do otherwise. Humans thus have a moral responsibility for their choices.

2.3 Freedom in Jesuit Letters

In the Jesuit Letters of February 9th, 1645, Descartes breaks more completely from the *Meditations* to affirm the importance of freedom in indifference. Commenting on the positive faculty of "The will," he says,

> I think it has it not only with respect to those actions to which it is not pushed by any evident reasons . . . but also with respect to all other actions; so that when a very evident reason moves us in one direction, although morally speaking we can hardly move in the contrary direction, absolutely speaking we can. For it is always open to us to hold back from pursuing a clearly known good, or from admitting a clearly perceived truth . . . (CSMK 245)

This view of indifference moves away from the position maintained in the *Meditations* and the *Principles* as it acknowledges its power even in the face of clear and distinct perceptions. To speak absolutely, humans have the freedom of indifference, or the "ability to do or not do something" in the words of the *Meditations*.

3. Responding to Descartes' Views of Freedom

I will now resituate Descartes' discussion of human freedom of epistemological pursuit within the field of theology. Under Descartes' theology, God the creator is also God the sustainer. He alone has endowed free will in men, while He also preordains all human actions and, in the *Meditations*, especially the inexorable acceptance of clear and distinct perceptions. This relates to the much-heated philosophical debate of compatibilism vs incompatibilism between libertarian free will and God's divine preordination, or theistic determinism. Therefore, it makes sense to first raise objections to Descartes using incompatibilist arguments then examine his texts on

our voluntary freedom in making non-predetermined judgments under an "Overdetermined Election" thought experiment.

3.1 Introduction to Incompatibilism

Compatibilism emerges as a response to the tension between libertarian free will and hard determinism, or in Descartes' specific case, theistic determinism. To explain their apparent contradiction, I will first present a summary of these two key concepts.

Conventionally, libertarians believe that free will is manifested in the Principle of Alternate Possibilities, the control a person has over her decision resulting from her ability to choose among alternative courses of actions. In our daily context, a famous architect creating his blueprint can design the building in many ways and therefore his choice to a large extent is free, while the construction workers must follow the singular, determined blueprint and are therefore less free. This ability to do otherwise is reflected in Descartes' work as the freedom of indifference.

On the other hand, the key claim of hard determinism is the future's predictability. It holds that the world is a deterministically predictable system (Stone 257). The combination of our present state of affairs with the set laws of nature can fully designate our future; the progression of events is based solely upon logic in nature. If everything is predestined this way, it is virtually impossible to do otherwise. Offering no alternative possibilities, this argument distrusts individuals' voluntary control over their actions and potentially threatens libertarian free will. For Descartes, this determinism presents itself as divine providence. As he states in the *Principles*, it is "certain that everything was preordained by God" (CSM I 206). God's all-determining power seems to preclude human free will, leading to incompatibilists' objections to Descartes.

3.2 Incompatibilist Objection to the *Meditations*

To delve deeper into Descartes' dismissal of freedom of indifference, or libertarian free will as defined above, in the *Meditations*, I will analyze one passage, dissecting his arguments on voluntarism. In the face of clear and distinct perceptions, "I could not," according to Descartes,

> but judge that something which I understood so clearly was true; but this was not because I was compelled so to judge by any external force, but because a great light in the intellect was followed by a great inclination in the will, and thus the spontaneity and freedom of my belief was all the greater in proportion to my lack of indifference." (CSM II 41)

This hints at the benefits of trading our inconsequential indifference for enhanced freedom of spontaneity. In this sense, certain knowledge, hampering freedom of indifference but boosting freedom of spontaneity, poses no threat to free will. But under the framework of libertarian free will, if our assent to ideas is predetermined by certain knowledge revealed to us by God, our choice has no real meaning. We know we must choose to accept certain facts even before our making the choice. As Descartes' definition of free will agrees with libertarianism (Ragland 379), this determinism seems to undermine epistemic freedom and contradicts himself. Are we truly free if we cannot do otherwise?

To meaningfully discuss this particular question, I will venture to present a

thought experiment. Suppose a fervent Democratic neurosurgeon inserts a magical chip inside each voter's head to secure a Democratic victory in the next election. The chip is designed to remain dormant and only activate when a voter decides to vote Republican, in which case it will impel her to vote Democrat. Now, suppose you were a voter who heads to the polls with the magical chip dormant inside your brain. You have been a committed Democrat and unquestionably want to vote Democrat. Without hesitating, you vote Democrat at the polling station and return home. Since your intention never deviated from voting Democrat, the device was never activated. However, if you had decided to vote Republican, you would have been prevented from doing so (Angelo 215). You are unfree according to the Principle of Alternate Possibilities because you could not have done otherwise. Once you are governed by an external chip, from that point on you stopped having the ability to make voluntary decisions. Even if you continue to act in a way you used to, your actions are now unfree because at the moment you lack the ability to renew your decisions.

This can help us rethink the unfair dismissal of indifference in the *Meditations*. In the context of our discussion on Descartes, "a great [divine] light in the intellect" installed in us by God is like the chip that predetermines our singular path of action. Our assent to clear and distinct perceptions is comparable to the act of voting Democrat in the thought experiment, while our impossible alternative of rejecting those perceptions is the equally impossible act of voting Republican. In the Meditations, as illustrated by the great light passage, Descartes' view of freedom dismisses the important ability to suspend judgment, analogous to voting Republican, when facing clear and distinct perceptions. Without this ability, it is as if we have taken a magical chip that predetermines our courses of action for us before we act. How can free will exist in such a case? Under the assault of incompatibilist arguments, Descartes fails to consistently demonstrate that free will and divine preordination are compatible.

3.3 Incompatibilist Objection to the *Principles*

In the *Principles*, Descartes outwardly admits his inability to reconcile human free will with God's preordination of certain knowledge. Although Descartes commits to both human freedom and divine providence, including predetermination of all events, as self-evident certainties, he recognizes an apparent conflict between these concepts: "We can easily get ourselves into great difficulties if we attempt to reconcile this divine preordination with our freedom of the will, or attempt to grasp both these things at once" (CSM I 206). Our attempt is especially undermined by our inability to demonstrate "how [divine power] leaves the free actions of men undetermined" (CSM I 206). This question raised by Descartes himself is so strong that his arguments cannot possibly hold together without giving legitimate answers. However, in response, he dismissed the question by asserting God's infinite perfection is beyond human understanding. This response is rather weak, especially placed in context with Descartes' attempts to attribute reasons to why God designed the world the way He did elsewhere. It seems like Descartes appeals to humans' limited understanding when it is useful for his arguments and appeals to reason when it is not. On a deeper level, Descartes fails to reconcile the constraint of divine providence with genuine human free will.

To return to the thought experiment, in this passage, as in the passage about automata from earlier in the Principles, determinism is a concern because Descartes' version of divine providence necessitates actions, the act of accepting certain

knowledge resembling the act of voting Democrat, to the point that they are no longer voluntary. God not only knew "from eternity whatever is or can be, but also willed it and preordained it" (CSM I 206). How can human freedom exist under this rigid framework? In the *Principles*, Descartes' claims appear weak under the challenge of libertarian arguments. There seems to be an obvious logical incoherence within Descartes' theist compatibilism between free will and determinism.

3.4 The Key to Descartes' Compatibilism

3.4.1 Libertarian Freedom in the Jesuit Letters

In the Jesuit Letters, Descartes seems to be denying that a great light in the intellect is followed by a great inclination in the will, maintaining that agents can "absolutely speaking" resist the determination of the will, even when "a very evident reason moves us in one direction," and even if "morally speaking we can hardly move on the contrary direction". In the context of the "Overdetermined Election" experiment, morality conditioned by society and the evident reason to give in to clear and distinct perceptions are comparable to a single entity: the magical chip's reinforcing influence. The key difference between these two scenarios lies in each cause's determinability. The thought experiment presents an example of overdetermination. Either the original intention or the magical chip alone is sufficient to determine the outcome of a Democratic vote. On the other hand, the two causes in Descartes' Jesuit passage closest to the magical chip, whether on their own or joined together, have no power to dictate a singular outcome of accepting clear and distinct perceptions; and the will, or the original intention, is not presented as predetermined and unchanging. The moral choice to believe in the evident reason and the 'immoral' choice to reject an evident reason are under the influence of the will at every moment. Thus, choices can be made and remade by a free agent at any point in time.

This difference in determinability also constitutes the fundamental difference in the discourse of free will between the Jesuit Letters and prior works. In the *Meditations* and the *Principles*, clear and distinct perceptions are portrayed as irrefutable by the will. According to the *Meditations*, "a great light in the intellect" is always necessarily followed by "a great inclination in the will" (CSM II 41). As the will must align with the intellect endowed in us by God, there is no freedom for us to decide against clear and distinct perceptions under the divine natural light. Equally, it is stated in the *Principles*, "We must believe everything which God has revealed, even though it may be beyond our grasp" (CSM I 203). The will guided by evident reason inevitably follows clear and distinct perceptions, and therefore the amalgamation around these clear and distinct perceptions is perfectly analogous to the all-determining magical chip. Free will in these two works is thus unlike the "absolute" freedom in the Jesuit Letters because it possesses no power to reject either evident reason or morality. By the Principle of Alternate Possibilities, aligning free will with a determined course of action completely constrains the will. All freedom without alternate possibilities, as in the "Overdetermined Election", are thus superficial. Only the circumstances in the Jesuit Letters differ fundamentally from our rigged election with a magical chip, and therefore only its version of

"absolute" freedom holds true under the test of alternate possibilities.

Moving beyond our thought experiment, we can find even more support for the Jesuit Letters over the two other texts in the real world. Practically, we see instances of Arkasia, the weakness of the will, in both ourselves and others. When evident reason or morality or even the two combined guide us in one direction, we feel a great motivation to act accordingly (Mele 116). However, perfect knowledge of the wrongness of certain actions does not guarantee we refrain from them. For example, there are 34.1 million smokers in the US (Petkovic). Exposed to anti-smoking campaigns, most of them understand smoking is costly and unhealthy, but against their better judgment, they still will to smoke and as a result smoke. In our everyday lives, to what extent is knowing the same as receiving the magical chip implant? Given the example of smoking, it must be admitted that having knowledge does not force knowers to follow any set course of action and thus is dissimilar to the magical chip in fundamental ways. Both the *Meditations* and the *Principles* are imprecise in their conception of the will, especially the overriding influence of its spontaneity. Alternatively, the Jesuit Letters provide a more credible view: "It is always open to us to hold back from pursuing a clearly known good, or from admitting a clearly perceived truth..." (CSMK 245). The active choice of the agent to refrain from what is good establishes a basis for moral responsibility. If crimes are predetermined, we have no reason to punish criminals who have no alternative choices but to offend. Only here in the Jesuit Letters can the rationale for punishment hold. Its conception of the will is not only freer by definition but also aligns better with real-world contexts.

3.4.2 Positioning Descartes in Compatiblism

Proceeding with the valid libertarian freedom in the Jesuit Letters, I attempt to reconcile human free will with divine providence in Descartes' works. Where exactly does Descartes fit under compatibilist arguments? The key to this puzzle seems to lie in Descartes' discussion of errors. In all of Descartes' mature works, such as the *Meditations* and the *Principles*, God is the creator who grants freedom and intellect to humans. When explaining the cause of errors, Descartes claims God "has given me the freedom to assent or not to assent in those cases..."; errors result in agencies who "misuse that freedom" (CSM II 42). In this situation, God is the source of all creaturely power, but the powers of creatures, even when efficaciously empowered by God, are really theirs, and so are distinct from His. If God efficaciously empowers me to assent to Descartes' argument, still the assenting is my action, not God's.

One way of expressing this difference might be as follows. While it seems clear that intramundane causation is transitive: if event A causes event B, and event B causes event C, then A causes C (Helm 117), due to the existence of "absolute" freedom as expressed in the Jesuit Letters, there is a distinct difference between granting agencies the power to perform actions and causing agencies to perform actions. According to the letter, our free action involves "a real and positive power to determine" that action (CSMK 234). The existence of libertarian free will lends a new perspective to think about this relationship. Say A represents God, B represents the agent's power to act, and C represents the agent's action. A grants B power to do C. Yet the agent has many choices and thus C has many alternatives. The result of empowerment in B can cause different configurations of C. A free agent can use B to

perform action C1 or action C2 . . . or action C100. God merely permits the agent to use power but does not decide each and every action. Therefore, in the unique case of divine willing permission, there is no necessary transitivity. This also helps to understand Descartes' initially perplexing claim that "neither divine grace nor natural knowledge ever diminishes freedom; on the contrary, they increase and strengthen it" (CSM II 40). With the help of certain knowledge, individuals can choose more wisely between alternative courses of action, although they need not choose wisely. With absolute freedom, their choosing wisely would be much more to their credit than cases in which they could not choose otherwise. Knowledge, as opposed to undermining free will and moral responsibility, rather enhances them. Thus, from my reading of Descartes' texts, neither divine knowledge nor learned intellect can undermine free will, which is by definition powerful and exists to retain alternative courses of action for each individual human.

4. Implications of Descartes' Epistemic Freedom

In the previous section, the paper resolved a central tension between knowledge and free will by the Principle of Alternate Possibilities. As long as free will exists within an individual in the first place, it cannot be undermined by the gaining of certain knowledge. However, philosophy is always open to discourse. This last section is dedicated to an exploration of the implications of the compatibilism proposed, raising more questions than answers.

4.1 Divine Freedom

Originally, if freedom is undermined by the possession of certain knowledge, the all-knowing God would be most unfree. With a compatibilism between certain divine knowledge and human free will, the conclusion seems to resolve the problem by induction and reassure us the all-knowing God can be free as well. However, we must carefully examine the alternative choices open to God by the Principle of Alternate Possibilities before we can form a conclusion. Returning to the sentence in the Jesuit Letters that epitomizes this requirement for alternatives, "For it is always open to us to hold back from pursuing a clearly known good, or from admitting a clearly perceived truth..." (CSMK 245), the requirement of free moral choice should be noted here. We must understand that Descartes' version of a supremely perfect God can choose neither between good and evil nor between truth and ignorance.

There is an example of how this lack of choice makes God unfree in part II of the *Principles*. As Descartes attempts to work out the principle around conservation of motion in the same amount, he cites divine immutability as evidence: "God's perfection involves not only his being immutable in himself, but also his operating in a manner that is always utterly constant and immutable" (CSM I 240). If this world of eternal immutability is the supreme perfection, is God free to create a less perfect world? Being all-good, He cannot purposefully do that. The more boundless His knowledge, the more unambiguous his path to perfection. Being all-knowing, He clearly has in mind what the best world looks like. Then, it seems that God has no choice but to be eternally perfect.

4.2 False Consciousness

In addition to God's lack of alternative choices, the text has another problem not discussed in the previous sections. Before we discuss the concept, let us first define false consciousness, a Marxist term that depicts proletariats as systematically brainwashed by capitalist ideologies (Meyerson 12). Outside the Marxist school of thought, false consciousness can denote people's inability to recognize wrongness in their society because of prevalent views that legitimize the existence of such wrongness.

My objection stems from Descartes' repeated emphasis on clear and distinct perceptions. Before, we have always taken these descriptions for granted. However, what are the criteria for "clear and distinct"? Is it how passionate you feel about an opinion? Then, ideologies ingrained in you are usually most passionately felt. Or is it something you are familiar with? Then, to only accept clear and distinct perceptions is to never advance into the unknown. In science as well as in diplomacy, the experimental attitude to navigate ambiguities is important, whether through trial and error or open negotiations. With only clear and distinct perceptions, we will never move beyond our present society's superstructure.

5. Conclusion

Descartes' strong commitment to both epistemic freedom and theist determinism is an interesting line to trace throughout his life's works. Exploring the philosopher's different conceptions of freedom in the Meditations, the Principles, and the Jesuit Letters, I argued the view of "absolute freedom" in the Jesuit Letters aligns best with true libertarian free will in the context of the compatibilism vs incompatibilism debate. Proceeding with the idea of "absolute freedom", the paper invokes the concept of transitivity to connect the dots between Descartes' seemingly inconsistent discussion of free will. When God permits our actions, he does not cause actions, and therefore our pursuit and acceptance of knowledge are not designated by God. By this point, I conclude that the possession of knowledge does not undermine free will. However, with this freedom comes moral responsibility. How do we assert our free will in the judgment given to us by our creator? Should we squander it by blindly embracing brainwashing ideologies? Or can we actively seek new perspectives to become truly open-minded? These are important questions to ask ourselves as we continue in our epistemological pursuits.

References

Angelo Corlett, J. "Moral Responsibility and History: Problems with Frankfurtian Nonhistoricism." *The Journal of Ethics*, vol. 22, no. 2, 2018, pp. 205-223.

Descartes, René, et al. *The Philosophical Writings of Descartes*. Cambridge [Cambridgeshire], Cambridge UP, 1984.

Helm, Paul. "God, Compatibilism, and the Authorship of Sin." *Religious Studies*, vol. 46, no. 1, 2010, pp. 115-124.

Kaufman, Dan. "Infimus Gradus Libertatis? Descartes on Indifference and Divine Freedom." *Religious Studies*, vol. 39, no. 4, 2003, pp. 391-406.

Mele, Alfred R. *Irrationality: An Essay on Akrasia, Self-deception, and Self-control.* plato.stanford.edu/entries/weakness-will/.

Meyerson, Denise. *False Consciousness*. Oxford, Clarendon Press, 1991.
Newman, Lex. "The Fourth Meditation." *Philosophy and Phenomenological Research*, vol. 59, no. 3, 1999, pp. 559-591.
Petkovic, Bojana. "27 Stunning Smoking Statistics and Facts." *Loud Cloud*, 13 Apr. 2022, loudcloudhealth.com/resources/smoking-statistics/.
Ragland, C. P. "Alternative Possibilities in Descartes's Fourth Meditation." *British Journal for the History of Philosophy*, vol. 14, no. 3, 2006, pp. 379-400.
Stone, J. "Free Will as a Gift from God: A New Compatibilism." *Philosophical Studies*, vol. 92, no. 3, 1998, pp. 257-281.

Theoretical Limitations of FTIR Spectroscopy

Ary Cheng

Author Background: *Ary Cheng grew up in the United States and currently attends BASIS Independent Silicon Valley in San Jose, California in the United States. His Pioneer research concentration was in the field of physics and titled "Fourier Series and Transforms with Applications in Physics and Related Fields."*

Abstract

Although Fourier Transform Infrared (FTIR) spectroscopy has proven to be the most successful method of infrared spectroscopy to date, it has its own set of limitations and drawbacks regarding spectral resolution. In this paper, the basic principles behind an FTIR spectrometer are explained, and the theoretical limitations of an FTIR spectrometer are analyzed quantitatively with the absorption spectra of carbon monoxide (CO) and water (H2O) in the range 3450 cm^{-1} – 4350 cm^{-1} (2300 nm – 2900 nm). An FTIR spectrometer is modeled mathematically by generating an interferogram of the original spectrum, then using the standard fast Fourier transform (FFT) algorithm in MATLAB to recover the spectrum. The sampling frequency of the interferogram and total mirror movement of the Michelson interferometer are varied from 1/2000 nm^{-1} to 1/10 nm^{-1} and from 0.1 cm to 50 cm respectively, and the effects on the resolution and accuracy of the recovered spectra are analyzed. A direct correlation between the total mirror movement and the spectral resolution is confirmed, while the sampling frequency of the interferogram is shown to have little effect on the recovered spectrum as long as it is at least twice the spatial frequency of the light in the region of interest. These findings are consistent with previous literature on this topic. Practical limitations and future directions in this field are also briefly discussed.

1. Introduction

Spectroscopy is one of the most efficient and accurate ways to determine an unknown substance's chemical composition to date. Unlike other methods that serve a similar purpose, spectroscopy can be performed on practically any sample, and does not fundamentally alter the sample in question [1]. When electromagnetic radiation from a broadband source passes through a sample, some of that radiation will be absorbed by the sample to take its molecules to a higher energy state, and the frequencies at which these absorptions occur is unique to each compound. Therefore, scientists can measure the intensity of the light passing through a substance over a wide range of frequencies to find the exact frequencies at which

this absorption occurs, and in turn, deduce the presence of certain molecules within an unknown sample. Almost all known compounds have characteristic absorption lines in the infrared (IR) region, making IR spectroscopy much more reliable and successful compared to other spectroscopic methods [2].

The origin of IR spectroscopy dates back to the late 19th century, with the invention of the bolometer by the American astronomer Samuel Langley in 1881 [3]. Even though Langley's bolometer was no more than a delicate and sensitive thermometer, he was able to measure the spectra of the Sun and the Moon as well as various other compounds in the IR region of the electromagnetic spectrum with his device. However, the bolometer quickly fell out of popularity as a device for infrared spectroscopy, and was eventually replaced by prism spectrometers, which were then replaced by dispersive spectrometers as high-quality diffraction gratings became easier and cheaper to manufacture in the 1960s [4]. The diffraction gratings allow only a small range of wavelengths to be measured by a detector at once, so accurate intensity measurements could be made for those wavelengths of light [5].

However, by the late 1980s, a new method for IR spectroscopy was introduced: Fourier Transform Infrared (FTIR) spectroscopy. As the name suggests, it utilizes a mathematical technique known as the Fourier transform to take measurements for all wavelengths of light simultaneously, instead of having to do it individually and separately as had been previously done with dispersive spectrometers. In particular, an FTIR spectrometer uses a Michelson interferometer to separate an incident beam of light into two different beams, which is then recombined to create an interference pattern. This interference pattern encodes information about all of the wavelengths of light at once, at which point a discrete Fourier transform (DFT) can be performed to return a spectrum that can be analyzed [6,7].

The FTIR method is both much faster than the traditional dispersion spectrometers and can provide spectra with higher resolution [8]. In addition to saving time by analyzing all wavelengths at once, it also bypasses the fundamental limitation with dispersion spectrometers, where a smaller range of wavelengths passing through each diffraction grating meant poorer spectral quality. However, the FTIR method is still far from perfect. There are a myriad of problems and limitations with this technique, which are briefly analyzed in this paper.

2. The FTIR Spectrometer

2.1. The Michelson Interferometer

One of the most crucial aspects of an FTIR spectrometer is the Michelson interferometer, which can collect information about a beam of light through its interference pattern with itself:

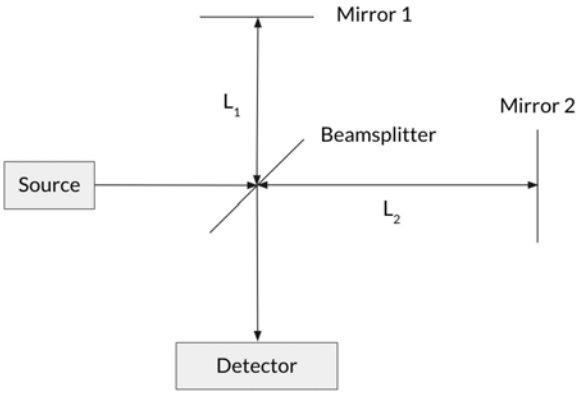


Figure 1. *A basic schematic diagram of a Michelson interferometer*

An incident beam of light is first split into two beams of equal intensity by a beamsplitter. Both beams are then reflected back using mirrors and recombined at the beamsplitter. Since both beams will have traveled different distances (2L1 and 2L2), they will be out of phase with each other, and interfere in a way that does not replicate the original beam of light. The resulting interference pattern is detected as an interferogram, which measures the intensity of the light as a function of the pathlength difference $\delta = 2(L2 - L1)$. The path-length difference is varied by controlling the movement of one mirror while keeping the other mirror stationary. Moving forward, the term "mirror movement" will refer to a change in path-length difference.

2.2. The Fourier transform applied to an interferogram

The Fourier transform, defined as follows [9],

$$F(s) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i s x} dx \tag{1}$$

turns a function of the variable x into that of a new variable, s. For any Fourier transform pair F(s) and f(x), the variables x and s must have inverse dimensionality; that is, the product xs must be dimensionless. Thus, the Fourier transform of an interferogram would be:

$$I(\tilde{v}) = \int_{-\infty}^{\infty} I(\delta) e^{-2\pi \tilde{v} \delta} d\delta \tag{2}$$

where $\tilde{v}$ is a quantity with units of inverse length. This quantity is known as spectroscopic wavenumber, and is simply defined as the reciprocal of wavelength [5].

Consider a monochromatic beam of light with wavenumber $\tilde{\nu}$. Light is an oscillation of electric and magnetic fields. Thus, right after it is split by the beamsplitter, the resulting light waves would have an electric field of $E_1 sin\left(2\pi\tilde{\nu}ct\right)$ and $E_2 sin\left(2\pi\tilde{\nu}ct\right)$, where E_1 and E_2 represent the amplitude (maximum magnitude) of the respective electric fields, *c* represents the speed of light, and *t* represents the time elapsed since the source began emitting light. After the beams of light travel through the arms of the interferometer and come back to the beamsplitter, they would have an electric field of $E_1 sin\left(2\pi\tilde{\nu}ct + 2\pi\tilde{\nu}d_1\right)$ and $E_2 sin\left(2\pi\tilde{\nu}ct + 2\pi\tilde{\nu}d_2\right)$ respectively, where d_1 and d_2 represent the total distances traveled by each beam of light through the arms of the interferometer. The recombined beam would then have an electric field E described by:

$$E = E_1 sin\left(2\pi\tilde{\nu}ct + 2\pi\tilde{\nu}d_1\right) + E_2 sin\left(2\pi\tilde{\nu}ct + 2\pi\tilde{\nu}d_2\right)$$

The intensity of a beam of light is proportional to the square of the field, so the recombined beam would have an intensity I given by:

$$I = E^2 = \left(E_1 sin\left(2\pi\tilde{\nu}ct + 2\pi\tilde{\nu}d_1\right) + E_2 sin\left(2\pi\tilde{\nu}ct + 2\pi\tilde{\nu}d_2\right)\right)^2$$
$$E_1^2 sin^2\left(2\pi\tilde{\nu}(ct + d_1)\right) + E_2^2 sin^2\left(2\pi\tilde{\nu}(ct + d_2)\right) + 2E_1E_2 sin$$

This expression can be simplified with the trigonometric identities $sin^2(x) = \frac{1-cos(2x)}{2}$ and $2sin(x)cos(x) = cos(x-y) - cos(x+y)$. This yields the following expression for I:

$$I = \frac{E_1^2}{2}\left[1 - cos\left(4\pi\tilde{\nu}(ct + d_1)\right)\right] + \frac{E_2^2}{2}\left[1 - cos\left(4\pi\tilde{\nu}(ct + d_2)\right)\right] + E_1E_2\left[cos\left(2\pi\tilde{\nu}\delta\right) - cos\left(2\pi\tilde{\nu}(d_1 + d_2 + 2ct)\right)\right]$$

where $\delta = d_1 - d_2$ is the path-length difference. Notice that the intensity changes with time. However, since light waves have an extremely high frequency, the detector can only measure the average intensity over time. Since all of the cosine terms containing t will oscillate between -1 and 1 over time, their average value is 0 and can therefore be ignored. This gives the following expression:

$$I = \frac{1}{2}(E_1^2 + E_2^2) + E_1E_2 cos\left(2\pi\tilde{\nu}\delta\right)$$

Now consider the original beam of light again. If it has an intensity of I_0 and an amplitude of E_0, then the two beams of light from the beamsplitter would have intensities of $I_1 = I_2 = I_0/2$. Since intensity is proportional to the square of the field, they would have amplitudes of $E_1 = E_2 = E_0/\sqrt{2}$. Thus we can rewrite *I* in terms of I_0:

$$I = \frac{1}{2}(E_1^2 + E_2^2) + E_1E_2cos(2\pi\tilde{\nu}\delta) = \frac{1}{2}\left(\frac{E_0^2}{2} + \frac{E_0^2}{2}\right) + \frac{E_0^2}{2}cos(2\pi\tilde{\nu}\delta) = \frac{I_0}{2} + \frac{I_0}{2}cos(2\pi\tilde{\nu}\delta)$$

Thus, if a monochromatic beam of light with wavenumber $\tilde{\nu}$ has an intensity of I_0 before hitting the beamsplitter of the Michelson interferometer, then its intensity I upon recombination at the beamsplitter can be described by [6,7,9,10]:

$$I = I_0 \cdot \frac{1 + cos(2\pi\tilde{\nu}\delta)}{2} \tag{3}$$

where δ represents the path-length difference between the two beams. For a full, continuous spectrum of light with intensity at wavenumber $\tilde{\nu}$ given by the function $I(\tilde{\nu})$, the total intensity I measured by the detector at a path-length difference δ would then be:

$$I(\delta) = \int_0^\infty I(\tilde{\nu}) \cdot \frac{1 + cos(2\pi\tilde{\nu}\delta)}{2} d\tilde{\nu} = \frac{1}{2}\int_0^\infty I(\tilde{\nu})d\tilde{\nu} + \frac{1}{2}\int_0^\infty I(\tilde{\nu})cos(2\pi\tilde{\nu}\delta)d\tilde{\nu} \tag{4}$$

Now consider the cosine form of the Fourier transform applied to intensity as a function of wavenumber:

$$I(\delta) = \int_{-\infty}^{\infty} I(\tilde{\nu})cos(2\pi\tilde{\nu}\delta)d\tilde{\nu} \tag{5}$$

Equation (4) and Eq. (5) only differ by a constant, and thus the intensity vs. wavenumber graph can be obtained through a Fourier transform of an interferogram [6,9]. Wavenumber can then be converted back into terms of wavelength if necessary to give an intensity vs. wavelength spectrum.

However, there is one final problem with this approach that must be resolved. The continuous Fourier transform, as defined in Eqs. (1) and (2), requires continuous data about light intensity, which would entail an infinitesimal change in δ, and, by extension, instruments capable of infinite precision. This is impossible to achieve in the real world. Instead, a discrete Fourier transform (DFT) is used, but this places a fundamental limitation on the spectra that can be recovered with this method. The Nyquist-Shannon theorem, also known as the sampling theorem, shows that the maximum frequency that can be recovered from a DFT is exactly half of the sampling frequency of the original data set [11]. In this context, the spatial frequency of a light wave is precisely equal to its wavenumber, so the sampling frequency of the interferogram should be at least twice the value of the highest wavenumber that needs to be recovered. For instance, a sampling rate of 1 sample per 100 nm (1/100 nm^{-1} = 100,000 cm^{-1}) for the interferogram is needed to recover information about light with wavenumbers less than 50,000 cm^{-1}.

3. Methodology

This paper uses a mathematical model of an FTIR spectrometer to analyze its limitations regarding resolution and accuracy. The radiation emitted by a broadband source in a physical FTIR spectrometer is continuous, but this will be impossible to model computationally, as an infinite number of data points would be required to simulate a true intensity distribution. Instead, this model will take one data point every 0.01 nm for the absorption spectrum. As a result, the interferogram will no longer be an integral spanning all possible wavelengths, as in Eq. (4), and it will instead be a discrete sum:

$$I(\delta) = \sum_j I(\tilde{v}_j) \cdot \frac{1 + cos(2\pi \tilde{v}_j \delta)}{2} \tag{6}$$

A derivation of this result by Prof. Frank Rioux of St. John's University can be found online [12]. It is largely similar to the derivation of the continuous integral version of this result, so it will not be shown here.

The interferogram is generated directly through Eq. (6), with no intermediate steps to simulate the actual interference pattern of light. The initial absorption spectrum and output spectrum are generated, graphed, and analyzed with MATLAB. The information for the absorption spectra used for the analysis (carbon monoxide and water) is taken from the HITRAN database [13]. The absorption spectra are calculated using a simplified version of the model presented in the LinePak™ library [14]. More information about the implementation of this model can be found in the Appendix.

This model will also use negative values for intensity in the interests of computational efficiency. Negative intensity has no physical meaning, but it can significantly reduce the amount of computation required to simulate the interferogram according to Eq. (6). If the background intensity at wavelengths where there are no spectral features is taken to be zero, then only the intensities in the region of interest would need to be included in the calculation, as values of zero elsewhere will not contribute to the sum. If the background intensity was instead chosen to be nonzero, then all wavelengths would need to be included in the calculation, significantly increasing the runtime of the simulation. Note that this is simply because the model presented here is purely mathematical, so the interferogram has to be generated mathematically as well. Real FTIR spectrometers do not need to work with negative intensities as they are merely measuring an interferogram instead of generating one.

3.1. An Exemplative Model

We begin with a simple, theoretical absorption spectrum as a conceptual demonstration. This spectrum is taken from 2200 nm to 2400 nm and consists of five absorption lines with wavelengths, wavenumbers, and relative intensities given in Table 1. Each absorption line is assumed to have an infinitesimal width.

Table 1. *Numerical values for the absorption lines of the idealized sample spectrum*

	Wavelength (nm)	Wavenumber (cm^{-1})	Relative Intensity (arb.)
Line 1	2250	4444.44	-0.25
Line 2	2300	4347.83	-0.50
Line 3	2320	4310.34	-0.20
Line 4	2330	4291.85	-0.80
Line 5	2380	4201.68	-0.40

Here are what the spectra and interferogram look like for these absorption features:

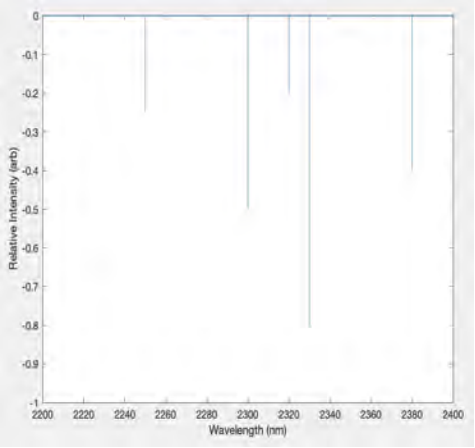


Figure 2. *Model spectrum (wavelength) with the absorption features in Table 1.*

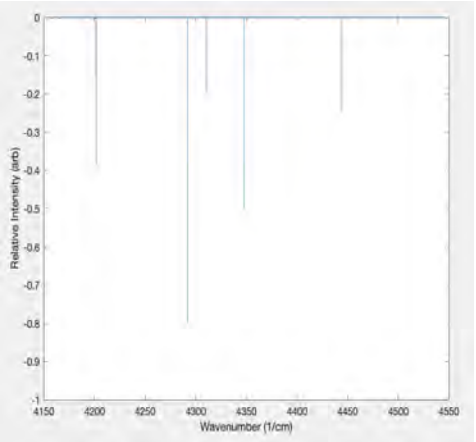


Figure 3. *Model spectrum (wavenumber) with the absorption features in Table 1.*

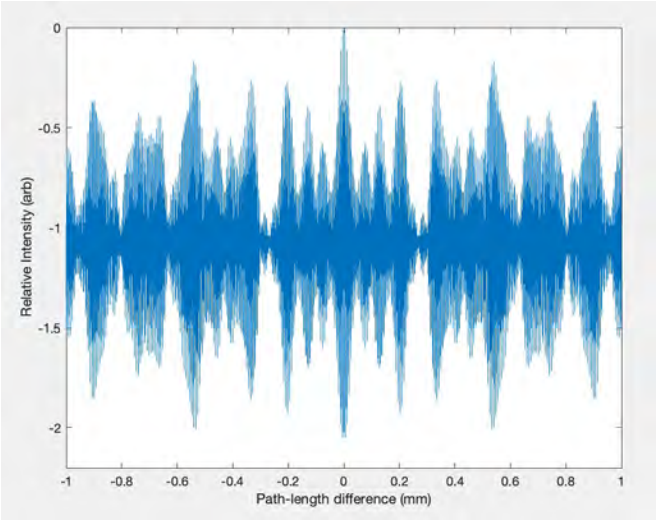


Figure 4. *Interferogram with sampling frequency 1/1000 nm^{-1}, or one sample every 1000 nm of mirror movement. Path-length difference ranges from -100 mm to 100 mm, but only -1 mm to 1 mm is shown here.*

Comparing Figs. (2) and (5) or Figs. (3) and (6), we can see that the FTIR spectrometer recovered the correct values of wavenumber/wavelength for all five absorption lines. However, the relative intensities for two of the absorption lines is slightly different from the expected value, and some of the output spectrum's absorption lines have a noticeable width near their "base" instead of effectively zero width. Section 4 is dedicated towards quantifying these effects for real absorption spectra. Analysis is only done on the intensity vs. wavenumber spectra going forward, since the spectra recovered through this method gives only an evenly spaced vector for wavenumbers.

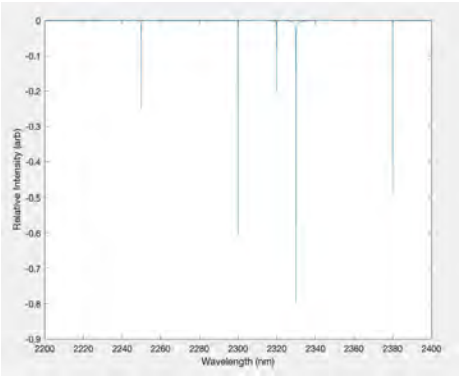


Figure 5. *Calculated output spectrum (wavelength) using the interferogram in Figure 4.*

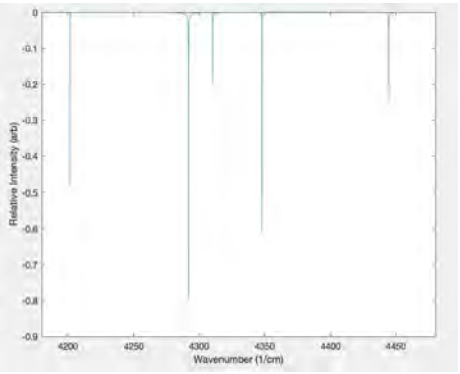


Figure 6. *Calculated output spectrum (wavenumber) using the interferogram in Figure 4.*

4. Data Analysis

There are two values that can be varied in the generation of each interferogram: total path-length difference and sampling frequency. Several values for these parameters are taken, and the spectra they recover are compared to the original spectrum. A quantitative comparison is difficult to conduct, as the input and output spectra usually have a different number of data points. Instead, the number and relative intensities of resolved spectral lines is compared.

4.1. Carbon Monoxide (CO)

One of the defining features of the IR spectrum of carbon monoxide is the first-overtone band, which occurs at around 2300 nm (4350 cm^{-1}) [15,16]. The wavenumber spectrum in this region is shown in Figure 7.

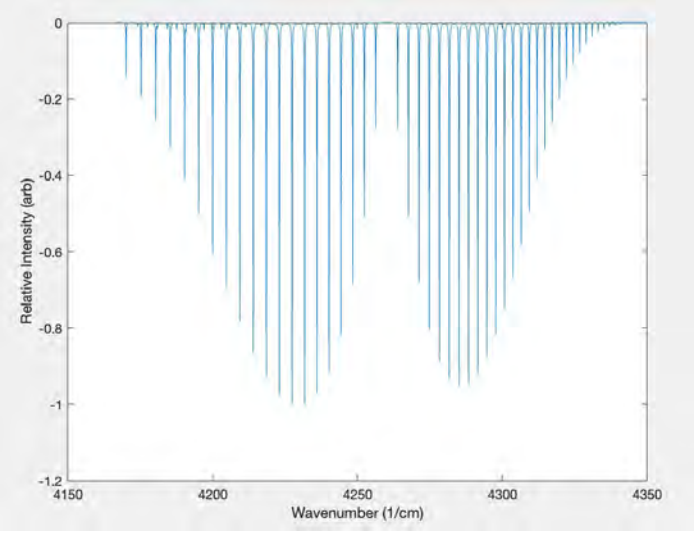


Figure 7. *Spectrum of CO from 4150 cm^{-1} to 4350 cm^{-1}. Data taken from Ref. (13).*

We begin by fixing the total mirror movement at 20 mm, so the path-length difference δ ranges from -10 mm to 10 mm. The sampling frequency of the interferogram (f_s) is varied, and the spectra recovered from four trials are shown in Figs. (8) – (11).

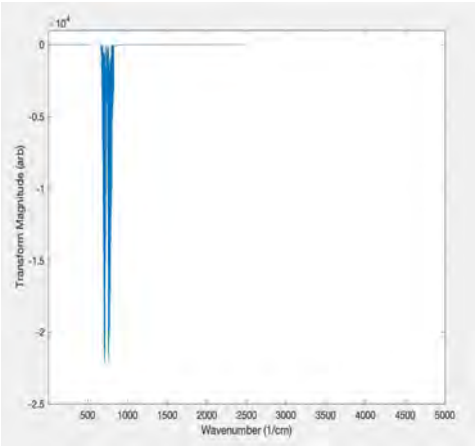


Figure 8. *Calculated spectrum of CO from 0 cm^{-1} to 5000 cm^{-1}. f_s = 1/2000 nm^{-1} = 5000 cm^{-1}, and δ is taken from -10 mm to 10 mm.*

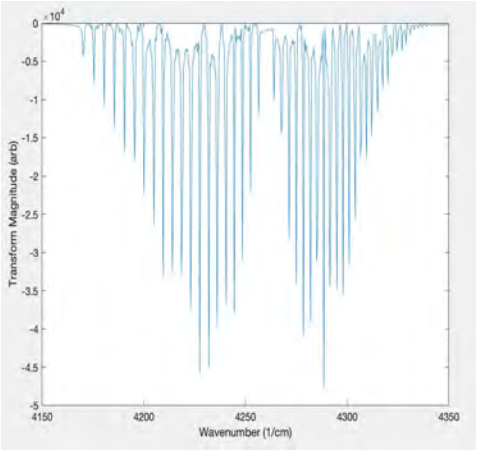


Figure 9. *Calculated spectrum of CO from 4150 cm^{-1} to 4350 cm^{-1}. f_s = 1/1000 nm^{-1} = 10000 cm^{-1}, and δ is taken from -10 mm to 10 mm.*

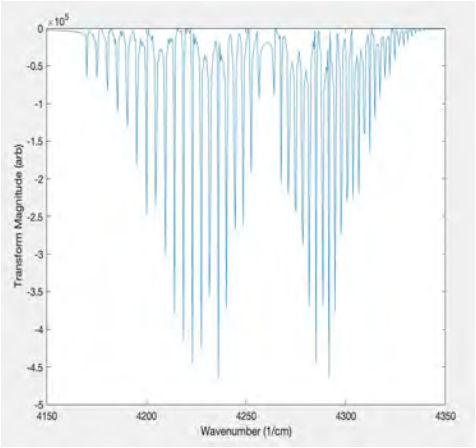


Figure 10. *Calculated spectrum of CO from 4150 cm^{-1} to 4350 cm^{-1}. f_s = 1/100 nm^{-1} = 1.0 $\cdot 10^5$ cm^{-1}, and δ is taken from -10 mm to 10 mm.*

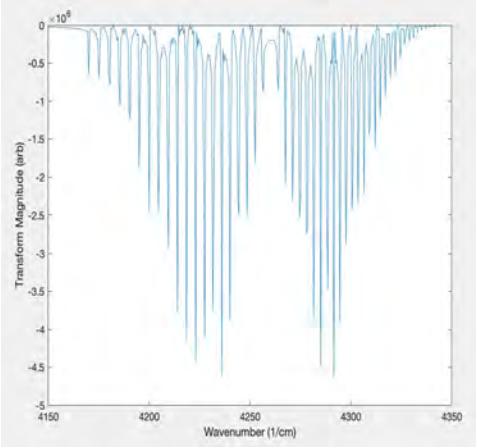


Figure 11. *Calculated spectrum of CO from 4150 cm^{-1} to 4350 cm^{-1}. f_s = 1/10 nm^{-1} = 1.0 $\cdot 10^6$ cm^{-1}, and δ is taken from -10 mm to 10 mm.*

Note the axes on Fig. (8). As described earlier, a fundamental limitation of a DFT is that it can only recover information about frequencies lower than the Nyquist frequency f_N, which is equal to $f_s/2$ for a sampling frequency of f_s. Thus, a sampling frequency of 1/2000 nm^{-1} = 5000 cm^{-1} can recover information about wavenumbers only up to 2500 cm^{-1}, which is not high enough to capture the spectral features around 4350 cm^{-1}. This is also why the absorption spectrum "stops" at precisely 2500 cm^{-1}. Interestingly, the same spectral features seem to show up at around 650 cm^{-1}. This is the product of aliasing, where signals for a frequency f_a above the Nyquist frequency $f_N = f_s/2$ show up at $f_{a'} = f_s - f_a$ [11]. In this case, since the sampling frequency is 5000 cm^{-1}, the spectral features around 4350 cm^{-1} will show up around 5000 cm^{-1} – 4350 cm^{-1} = 650 cm^{-1}.

The other three sampling frequencies are high enough to resolve the spectral features at 4350 cm^{-1}, but a comparison with the original spectrum (Figure 7) shows that they are still somewhat inaccurate, as can be seen by the widened "bases" of the absorption lines and the noticeably incorrect relative intensity for some of the lines, particularly around 4220 cm^{-1} and 4275 cm^{-1}. Additionally, a higher sampling frequency past the amount needed to resolve the spectral features seems to have no effect on the accuracy of the output, as there is almost no noticeable difference between Figure 10 and Figure 11. However, the reason for the relatively low accuracy can be easily seen when a small region of the original and output spectra are compared.

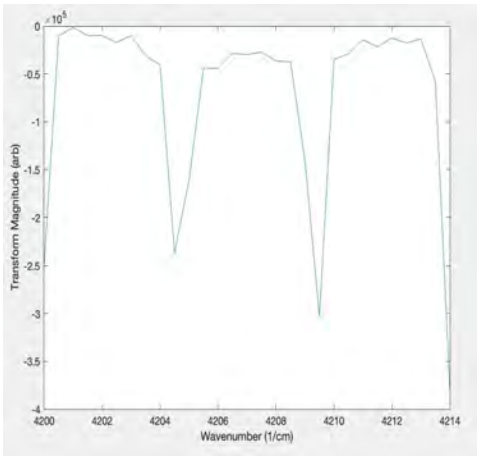


Figure 12. *Calculated spectrum of CO from 4200 cm^{-1} to 4214 cm^{-1}. f_s = 1/100 nm^{-1}, and δ is taken from -10 mm to 10 mm.*

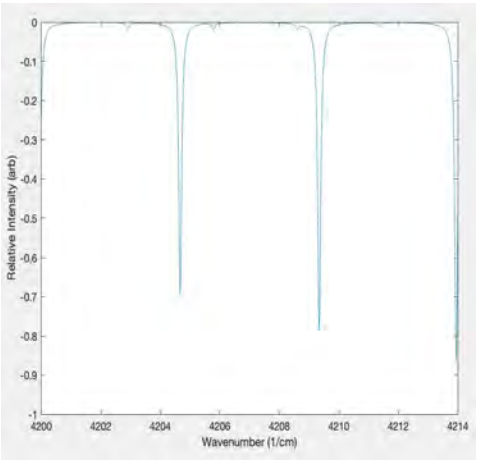


Figure 13. *Original spectrum of CO from 4200 cm^{-1} to 4214 cm^{-1}.*

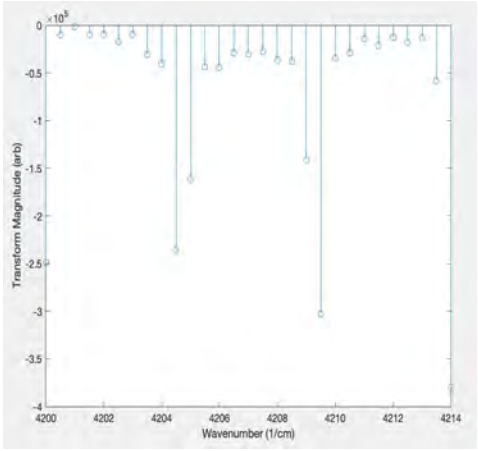


Figure 14. *Stem plot of the calculated spectrum of CO from 4200 cm^{-1} to 4214 cm^{-1}. There are a total of 29 data points for a spectral resolution of 0.48 cm^{-1}.*

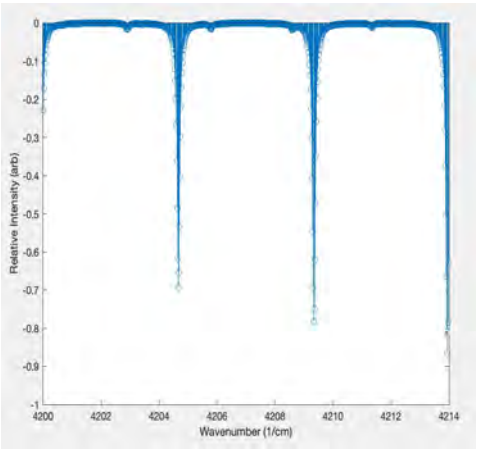


Figure 15. *Stem plot of the original spectrum of CO from 4200 cm^{-1} to 4214 cm^{-1}. There are a total of 788 data points for a spectral resolution of 0.0178 cm^{-1}.*

The output spectrum is clearly degraded because it does not have a high enough resolution, with only 29 data points in this region. This value is the same for all of the three sampling frequencies shown in Figures 9 – 11, which is consistent with the quick, qualitative analysis done prior to Figure 12. This gives a spectral resolution of (4214 cm^{-1} – 4200 cm^{-1})/29 = 0.48 cm^{-1}, meaning that only spectral features wider than 0.48 cm^{-1} apart can be distinguished. A stem plot of both the original and observed spectra in this region is shown in Figures 14 and 15 for a clearer illustration of the consequences of such a low spectral resolution. By comparison, the original spectrum has nearly 800 data points in the same region, allowing it to easily resolve the less noticeable spectral features at 4206 cm^{-1} and 4208.5 cm^{-1}. These are actually absorption lines that are much lower in intensity, but the low resolution of the output spectra makes them impossible to recognize without prior knowledge of their existence, as is often the case when performing spectroscopy on an unknown sample.

The remainder of this section (4.1) focuses on the effects of varying the amount of total mirror movement. Recall that the path-length difference between the two beams of light in a Michelson interferometer is varied by moving one of the two mirrors. Thus, the range of path-length difference values is synonymous with the amount of total mirror movement, and these two terms will be used interchangeably in the rest of this section. Since sampling frequency has no effect on the resolution of the output spectrum, it is kept constant at 1/100 nm^{-1} = 100,000 cm^{-1} for the following trials, and the total mirror movement is varied.

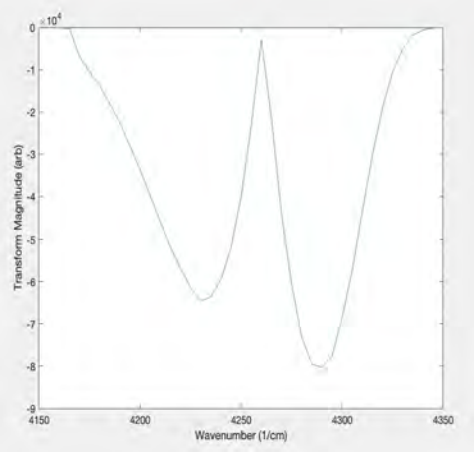


Figure 16. *Calculated spectrum for CO from 4150 cm^{-1} to 4350 cm^{-1}. f_s = 1/100 nm^{-1} = 100,000 cm^{-1}. Path-length difference is varied from -0.5 mm to 0.5 mm (0.1 cm total mirror movement). Spectral resolution is calculated to be 10 cm^{-1}.*

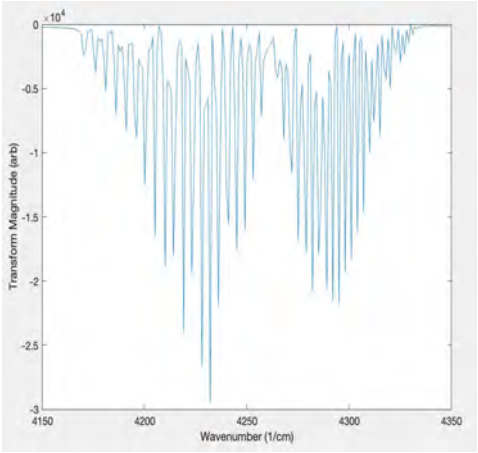


Figure 17. *Calculated spectrum for CO from 4150 cm^{-1} to 4350 cm^{-1}. f_s = 1/100 nm^{-1} = 100,000 cm^{-1}. Path-length difference is varied from -5 mm to 5 mm (1 cm total mirror movement). Spectral resolution is calculated to be 1 cm^{-1}.*

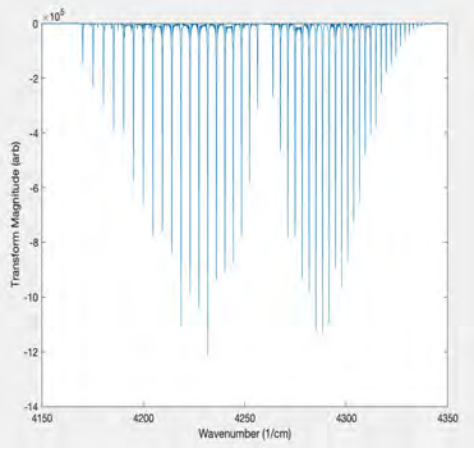


Figure 18. *Calculated spectrum for CO from 4150 cm^{-1} to 4350 cm^{-1}. f_s = 1/100 nm^{-1} = 100,000 cm^{-1}. Path-length difference is varied from -50 mm to 50 mm (10 cm total mirror movement). Spectral resolution is calculated to be 0.1 cm^{-1}.*

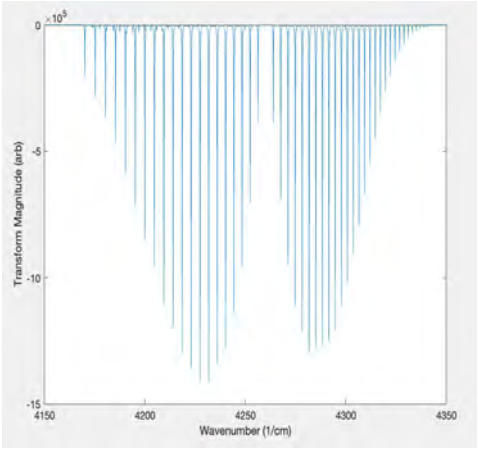


Figure 19. *Calculated spectrum for CO from 4150 cm^{-1} to 4350 cm^{-1}. f_s = 1/100 nm^{-1} = 100,000 cm^{-1}. Path-length difference is varied from -250 mm to 250 mm (50 cm total mirror movement). Spectral resolution is calculated to be 0.02 cm^{-1}.*

There is a clear increase in both resolution and accuracy as the total mirror movement is increased. Figure 19 and Figure 7 are nearly identical except for a few intensity values around 4280 cm^{-1}. Again, a small region of the spectra (4200 cm^{-1} to 4214 cm^{-1}) is analyzed using a stem plot as shown in Figure 20.

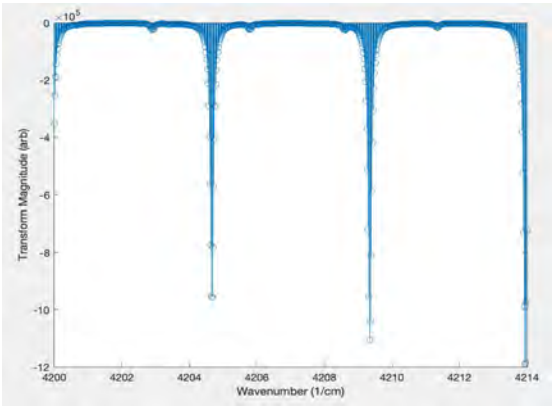


Figure 20. *Stem plot of the calculated spectrum of CO from 4200 cm^{-1} to 4214 cm^{-1}. There are a total of 700 data points for a spectral resolution of 0.02 cm^{-1}.*

Across this region, this output spectrum has a total of 700 data points, giving it a spectral resolution of 0.02 cm^{-1}. It perfectly captured the small spectral features that were present in Figure 15. By comparison, the original spectrum has a spectral resolution of around 0.0178 cm^{-1}, explaining why the two spectra were so similar in terms of resolution and accuracy.

4.2. Water (H2O)

A similar analysis is performed on the absorption spectrum of water around 2500 nm (4000 cm^{-1}), where it has a group of absorption lines similar in intensity to the CO bandhead at 2300 nm [17]. However, unlike the relatively distinct and separate lines in the CO spectrum, water has many more absorption lines across this range and most absorption lines will overlap in a spectrum shown in Figure 21.

Similar to the analysis of CO in section 4.1, this section starts by keeping the total mirror movement constant at 20 mm (2 cm), and verifies the conclusion about the sampling frequency's effect on the recovered spectrum.

As expected, wavenumbers above the Nyquist frequency of $f_N = f_s/2 = 2500$ cm^{-1} were not recovered in Figure 22, and the spectral features are instead aliased at around 1500 cm^{-1}. Additionally, there is no noticeable difference between Figures 23 – 25, consistent with the analysis for the CO spectrum as well. Now keeping the sampling frequency constant at 1/100 nm^{-1} = 100,000 cm^{-1}, the total mirror movement is varied, as done for the CO spectrum. However, since the absorption spectrum for water has so many overlapping features, a qualitative analysis across the entire region is impossible. Instead, only a small region (3920 cm^{-1} to 3955 cm^{-1}) is analyzed, in similar fashion to Figures 12 – 15.

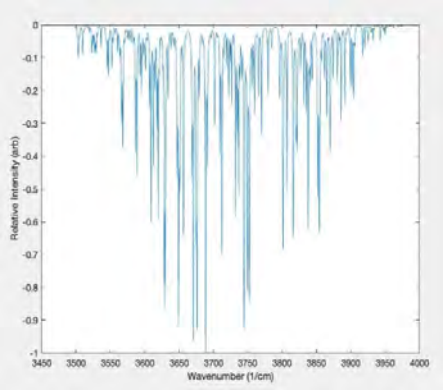


Figure 21. *Spectrum of H_2O from 3450 cm^{-1} to 4000 cm^{-1}. Data taken from Ref. [13].*

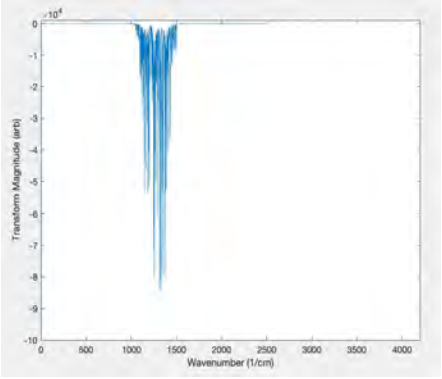


Figure 22. *Calculated spectrum of H_2O from 0 cm^{-1} to 4000 cm^{-1}. f_s = 1/2000 nm^{-1} = 5000 cm^{-1}, and δ is taken from -10 mm to 10 mm.*

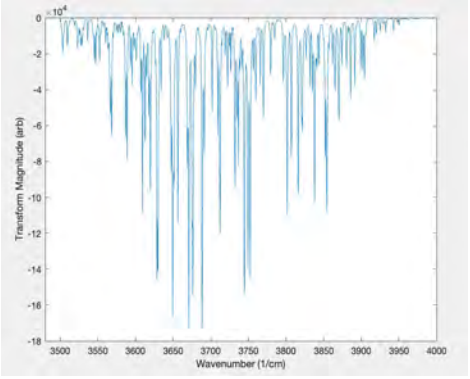


Figure 23. *Calculated spectrum of H_2O from 3500 cm^{-1} to 4000 cm^{-1}. f_s = 1/1000 nm^{-1} = 10000 cm^{-1}, and δ is taken from -10 mm to 10 mm.*

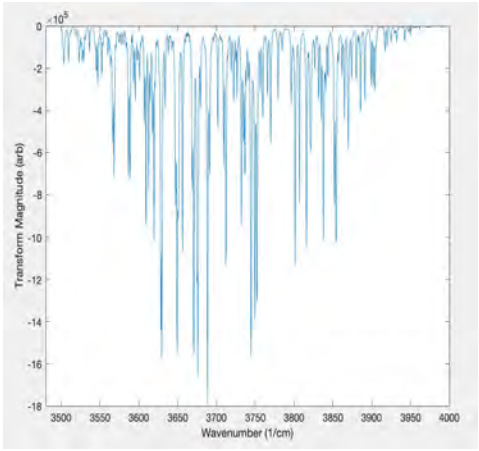


Figure 24. *Calculated spectrum of H_2O from 3500 cm^{-1} to 4000 cm^{-1}. f_s = 1/100 nm^{-1} = 1.0 ·10^5 cm^{-1}, and δ is taken from -10 mm to 10 mm.*

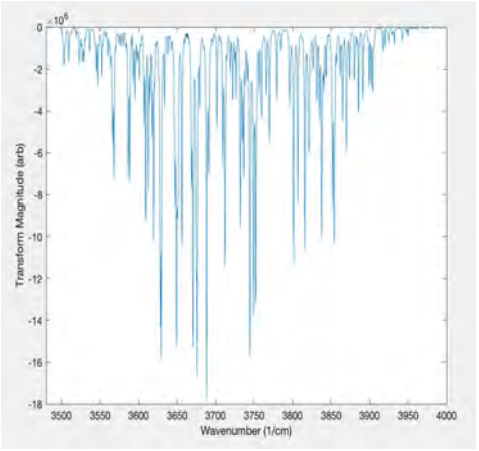


Figure 25. *Calculated spectrum of H_2O from 3500 cm^{-1} to 4000 cm^{-1}. f_s = 1/10 nm^{-1} = 1.0 ·10^6 cm^{-1}, and δ is taken from -10 mm to 10 mm.*

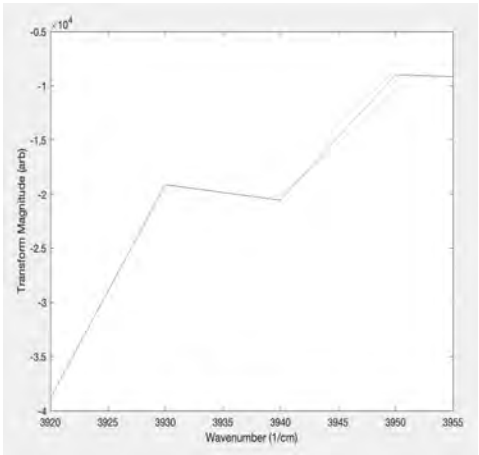


Figure 26. *Calculated spectrum for H_2O from 3920 cm^{-1} to 3955 cm^{-1}. f_s = 1/100 nm^{-1} = 100,000 cm^{-1}. Path-length difference is varied from -0.5 mm to 0.5 mm (0.1 cm total mirror movement). Spectral resolution is calculated to be 10 cm^{-1}.*

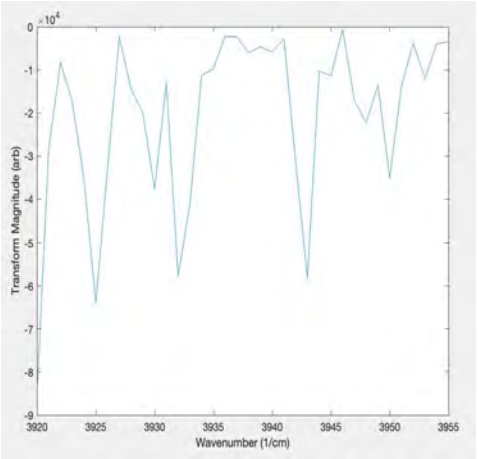


Figure 27. *Calculated spectrum for H_2O from 3920 cm^{-1} to 3955 cm^{-1}. f_s = 1/100 nm^{-1} = 100,000 cm^{-1}. Path-length difference is varied from -5 mm to 5 mm (1 cm total mirror movement). Spectral resolution is calculated to be 1 cm^{-1}.*

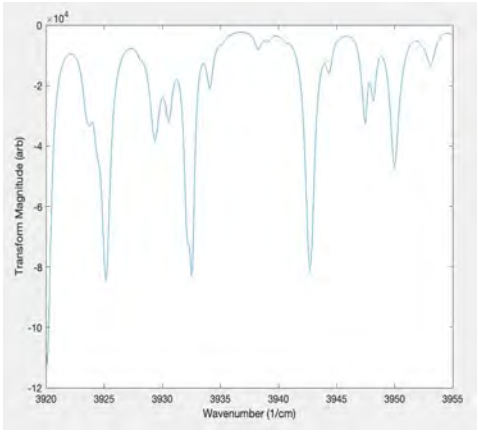


Figure 28. *Calculated spectrum for H_2O from 3920 cm^{-1} to 3955 cm^{-1}. f_s = 1/100 nm^{-1} = 100,000 cm^{-1}. Path-length difference is varied from -50 mm to 50 mm (10 cm total mirror movement). Spectral resolution is calculated to be 0.1 cm^{-1}.*

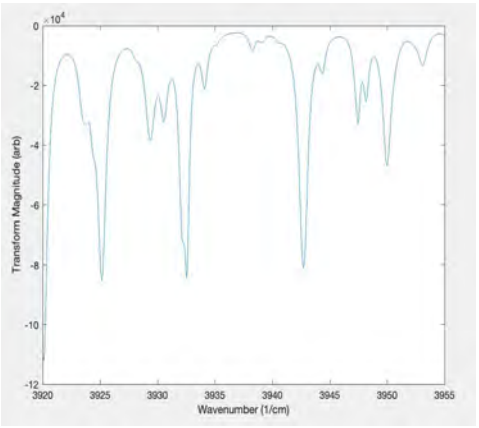


Figure 29. *Calculated spectrum for H_2O from 3920 cm^{-1} to 3955 cm^{-1}. f_s = 1/100 nm^{-1} = 100,000 cm^{-1}. Path-length difference is varied from -250 mm to 250 mm (50 cm total mirror movement). Spectral resolution is calculated to be 0.02 cm^{-1}.*

As expected, a larger value for total mirror movement gave a spectrum of higher resolution and accuracy, with Figure 28 and 29 almost looking identical except at the "tips" of their absorption lines. Comparing Figure 29 to the actual spectrum in this region shown in Figure 30, the graphs look identical.

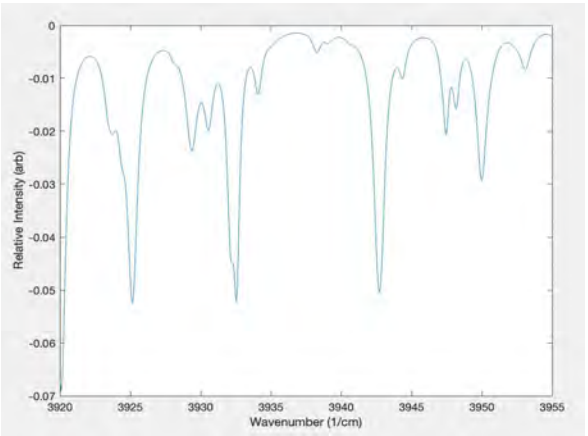


Figure 30. *Spectrum for H_2O from 3920 cm^{-1} to 3955 cm^{-1}. Data taken from Ref. [13].*

Now we compare the spectral resolutions of the two spectra. For the recovered spectra, a total of 1750 data points across a range of 35 cm^{-1} gives a spectral resolution of 0.02 cm^{-1}. The original spectrum has a similar spectral resolution of 0.0155 cm^{-1} with 2259 data points.

5. Limitations of the FTIR Spectrometer

5.1. Theoretical Limitations

The analysis in Section 4 demonstrates the effects of the sampling frequency of the interferometer and the total mirror movement on the recovered spectrum. In particular, the sampling frequency needs to be at least twice the spatial frequency of the light in the region of interest, but increasing frequency beyond this threshold has no significant effect. Additionally, a larger total mirror movement will increase both the resolution and accuracy of the recovered spectrum.

The analysis also demonstrates the existence of a theoretical limit of the resolution of an FTIR spectrometer, which is dependent on the maximum path-length difference L of the two light beams. This limit R can be quantitatively derived and shown to be [7]:

$$R = \frac{1}{2L} (7) \tag{7}$$

This result is consistent with the analysis in Section 4; the spectral resolutions of Figures 16 – 19 and Figures 26 – 29 perfectly match the value predicted by Eq. (7). Note that even though the model presented here takes a range of path-length difference centered around 0, the spectral resolution only depends on the maximum path-length difference, not the range. Thus, a range of -250 mm to 250 mm of path-length difference, as in Figure 19 and 29, will have a value of 250 mm for L.

There is another assumption made by this model; namely, that all rays of light from the beamsplitter will hit the mirrors at an angle of exactly 90◦. While this is theoretically true for a mirror that is an infinite distance away from the beamsplitter, real-world FTIR spectrometers are not of infinite length, which leads to a slight modulation in the interferogram. This effect can affect both the intensity and the perceived wavenumber of light [7]. As a result, most FTIR interferometers in the real world will multiply the interferogram by an apodization function to counteract this effect, with the most common apodization function for a path-length difference δ being [4]:

$$f(\delta) = \frac{1 + cos(\pi\delta)}{2} \tag{8}$$

In general, apodization functions have a negative impact on the spectral resolution, and Eq. (8) is a good general purpose apodization function since it achieves the desired results without significantly altering the spectral resolution. More complex functions are needed to better preserve the spectral resolution.

5.2. Practical Limitations

In addition to the theoretical limitations presented above, an FTIR spectrometer in the real world also faces physical challenges. As with all instruments in the real world, all signals received by an FTIR spectrometer contain some levels of noise, which can come from a myriad of different sources such as the detection of

background blackbody radiation, or simply the scattering of light as it travels along the two arms of the Michelson interferometer [7]. Additionally, a "perfect" FTIR spectrometer would need a beamsplitter capable of perfectly splitting a beam of light into two without any loss in intensity, and mirrors capable of perfectly reflecting a beam of light back towards the beamsplitter without any absorption. In the real world, this is impossible. There will always be a source of error that would decrease the accuracy of the spectrum recovered by an FTIR spectometer, and a perfect recovery of spectral features can never be achieved by any FTIR spectrometer as a result.

6. Conclusion

There are many applications of FTIR spectroscopy, ranging from identification of chemical molecules [18,19] to determining energy transitions associated with molecular vibrations [20,21,22]. For the former purpose, FTIR spectrometers often only need to have spectral resolution of 100 cm^{-1} or higher [8] (here, "higher" refers to a smaller value for spectral resolution), while FTIR spectrometers for the latter purpose can often have spectral resolutions as high as 0.00096 cm^{-1} [22]. As shown by the data analysis and Eq. (7), a spectral resolution this high would require a maximum path-length difference of more than 5 m, and this value will only increase for spectra with even better resolution. However, this is extremely difficult to achieve in practice, as moving the mirror across such a large distance will almost certainly cause it to tilt with respect to the beam of light, which will change the effective path-length difference [8]. Other effects such as light scattering also become much more prominent at such long distances.

One way of miniaturizing and optimizing the FTIR spectrometer is with the use of micro-electromechanical systems (MEMS) [8]. This is done by increasing the travel range of the mirrors while decreasing the size of the overall apparatus. A recent design introduced in 2010 allowed for a maximum path-length difference of 1 mm with an interferometer with dimensions of 35 mm × 35 mm × 65 mm [23]. Other designs utilize an interferometer known as the Mach-Zehnder interferometer, which operates on the same principles as the Michelson interferometer but produces two transmission outputs, which decreases the overall noise in the detector's measurements [8].

The mathematics behind the FTIR spectrometer also explains why it is the preferred method for infrared spectroscopy as opposed to visible or ultraviolet (UV) spectroscopy. The highest frequency of light that can be resolved directly depends on the sampling frequency of the interferogram, which can be measured in terms of time or distance depending on the type of instrument used. Some FTIR spectrometers move the mirrors at a constant velocity, while the dectector would take measurements at set time intervals, effectively taking samples at different values of path-length difference [19]. Another method is to directly move the mirror at the desired step-size, and technology today allows for stepsizes of around 80 nm with the use of piezoelectric instruments, corresponding to a step-size of 160 nm for path-length difference and a maximum observable wavenumber of 31250 cm^{-1}, or a wavelength of 320 nm [24]. This is barely shorter than the wavelength of visible violet light (380 nm), explaining why Fourier transform spectroscopy is predominantly used in the IR region of a spectrum.

7. Appendix: Generating Spectra

A line-by-line model, the most accurate way to model molecular absorption spectra, is simply the product of all of the individual absorption lines. Although each line corresponds to one energy transition, the change in energy associated with this transition varies and is measured as a broadened spectral line. For samples on Earth, there are two dominant effects that can affect each absorption line:

7.1. Doppler Broadening

All particles above the temperature of absolute zero will have some random thermal motion as a result of that temperature. Recall that for a light source moving along an observer's line of sight, the light will experience a Doppler shift and be measured at a different wavelength. The velocity of gas particles follows a Maxwell-Boltzmann distribution, and the cumulative effect of the Doppler shifts across all of the particles can be shown to be a Gaussian distribution around the original wavelength λ_0. The full-width at half-maximum (FWHM) and standard deviation σ of this distribution at a temperature T can be shown to be [25]:

$$FWHM = 2\lambda_0\sqrt{2ln2\frac{kT}{mc^2}} \tag{9}$$

$$\sigma = \frac{FWHM}{2\sqrt{2ln2}} = \lambda_0\sqrt{\frac{kT}{mc^2}} \tag{10}$$

where k represents the Boltzmann constant, *m* represents the mass of the particle, and *c* represents the speed of light. For the purposes of this paper, *T* is taken to be 300 K to match atmospheric conditions.

7.2. Pressure (Collision) Broadening

Molecular collisions can disturb the molecules' energy states and create a wide range of energy transitions, which is then observed as a broadened absorption line. Although this broadening always takes the form of a Lorentzian distribution, the specific parameters are impossible to calculate theoretically and must be determined experimentally. These parameters are given at pressure = 1 atm and temperature = 296 K in the HITRAN database [13], and since they closely match atmospheric conditions on Earth, the values will be used without further correction. In reality, the amount of pressure broadening can vary wildly for different values of pressure, and an example is given in Figure 31.

Additionally, the amount of pressure broadening for an absorption line also depends on the identity of the molecules in the collision. For instance, two molecules of carbon dioxide (CO2) colliding with each other would have a larger effect on the CO2 absorption spectrum than a collision between a carbon dioxide molecule and a molecule of oxygen. HITRAN provides the broadening parameters for both of these

cases, but only the self-broadening parameters will be considered, as all samples are assumed to be pure for the purposes of this paper.

Each absorption line experiences both of the broadening effects outlined above, so the true spectral distribution for each absorption line is a convolution of the Gaussian distribution due to Doppler broadening and the Lorentzian distribution due to pressure broadening. The resulting spectral profile is then scaled by the relative intensity of that particular absorption line. All spectral profiles are then multiplied together to return the full absorption spectrum.

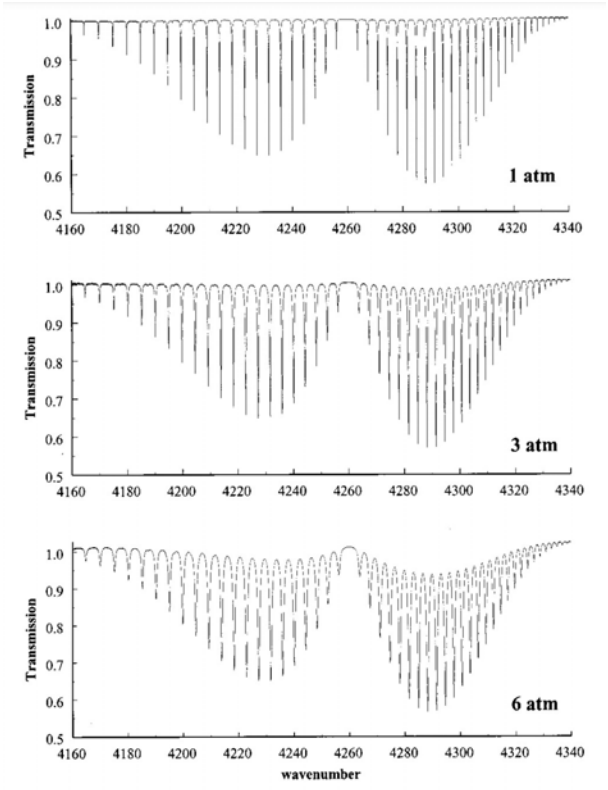


Figure 31. *The first-overtone band of the CO spectrum with different values for pressure. It is clear that higher pressures lead to broader spectral lines. Picture taken from Ref. [16].*

References

[1] B.H. Stuart, *Infrared Spectroscopy: Fundamentals and Applications*. (John Wiley & Sons, Chichester, UK, 2004). pp. 1-2.
[2] Z. Bacsik, J. Mink, and G. Keresztury, Applied Spectroscopy Reviews **39** (3), 295-363 (2004).
[3] R.B. Barnes and L.G. Bonner, American Journal of Physics **4**, 181-189 (1936).
[4] B.H. Stuart, *ibid.*, pp. 16-21.
[5] P. Larkin, *IR and Raman Spectroscopy*. (Elsevier, Oxford, UK, 2011). pp. 27-28.
[6] D. Pengra, (2017). Michelson Interferometer Fourier Transform Spectrometry. http://courses.washington.edu/phys331/michelson/ michelson.pdf. Accessed 15 Aug. 2022.
[7] D.R. Hearn, (1999). Fourier Transform Interferometry, Technical Report 1053, https://apps.dtic.mil/sti/pdfs/ADA370423.pdf. Accessed 15 Aug. 2022.
[8] J. Chai et al., Micromachines **11** (2), 214 (2020). DOI: https://doi.org/10.3390/mi11020214
[9] J. F. James, *A Student's Guide to Fourier Transforms with Applications in Physics and Engineering*, 3rd ed. (Cambridge University Press, Cambridge, 2011). pp. 8-11.
[10] *Laser Interferometry*. http://phyweb.physics.nus.edu.sg/L3000/Level3manuals/Laser%20Interferometry.pdf. Accessed 6 Sept. 2022.
[11] J. F. James, *ibid.*, pp. 33-35.
[12] F. Rioux. (2020). 211: The Michelson Interferometer and Fourier Transform Spectroscopy.
https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Quantum_Tutorials_(Rioux)/Spectroscopy/211%3A_The_Michelson_Interferometer_and_Fourier_Transform _Spectroscopy. Accessed 15 Aug. 2022.
[13] L.S. Rothman et al., Journal of Quantitative Spectroscopy and Radiative Transfer **110** (9-10), 533572 (2009). DOI: https://doi.org/10.1016/j.jqsrt.2009.02.013
[14] L.L. Gordley, B.T. Marshall, and D.A. Chu, Journal of Quantitative Spectroscopy and Radiative Transfer **52** (5), 563-580 (1994). DOI: https://doi.org/10.1016/0022-4073(94)90025-6
[15] G. Li et al., Applied Physics B **119**, 287-296 (2015). DOI: https://doi.org/10.1007/s00340-0156056-6
[16] A. Predoi-Crossa et al., The Journal of Chemical Physics **113**, 158 (2000). DOI: https://doi.org/10.1063/1.481783
[17] N. L. Kazanskiy, S. N. Khonina, and M. A. Butt, Photonics **9** (5), 331 (2022). DOI: https://doi.org/10.3390/photonics9050331
[18] E. N. Lewis, L. H. Kidder, and I. W. Levin, Microscopy and Microanalysis **3** (S2), 831-832 (2020). DOI: https://doi.org/10.1017/S1431927600011041
[19] C. Ruckebusch et al., Vibrational Spectroscopy **35** (1-2), 21-26 (2004). DOI: https://doi.org/10.1016/j.vibspec.2003.11.002
[20] T.L. Tan et al., Journal of Molecular Spectroscopy **321**, 59-62 (2016). DOI: https://doi.org/10.1016/j.jms.2016.02.003

[21] A. Batra et al., Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy **51** (1), 71-77 (1995). DOI: https://doi.org/10.1016/0584-8539(94)E0077-N
[22] B. Amyay et al., Chemical Physics Letters **491**, 17-19 (2010). DOI: http://dx.doi.org/10.1016/j.cplett.2010.03.053
[23] F. Merenda et al., Proceedings of SPIE **7680**, 78600V (2010). DOI: https://doi.org/10.1117/12.849670
[24] D. Michal, S. Matus, and M. Stefan, Review of Scientific Instruments **91**, 033102 (2020). DOI: https://doi.org/10.1063/1.5119206
[25] R. Nave. Broadening of Spectral Lines. http://hyperphysics.phy-astr.gsu.edu/hbase/Atomic/ broaden.html.
Accessed 20 Aug. 2022.

A Qualitative Analysis of Social Mobility, Financial Inheritance, and Wealth Accumulation within Black Households

Julius Dorsey

***Author background**: Julius Dorsey grew up in the United States and currently attends Regis High School in New York City in the United States. His Pioneer research concentration was in the field of political science and titled "Race, Religion, and Politics."*

Abstract

The intersection of race and class has historically divided America's political, economic, and social institutions. From race-based economic discrimination to modern-day racial wealth gaps, systemic racism has barred Black families in the United States from climbing the upward social mobility ladder. The crux of this economic and social inequality lies in the intergenerational transmission of crime, pain, poverty, segregation, and inequity. These systems have transcended generations of Black households, meaning that one generation can inherit the structures that oppressed the previous generation. Alternatively, social and economic progress—the antithesis of social and economic inequality—also transcends generations. Families with an abundance of wealth can nurture their young with the social and cultural capital necessary to maintain the wealth.

To bolster the collective understanding of how social progress and inequality live throughout generations, I documented the life events of Black families by conducting a series of interviews. A qualitative analysis of their testimonies unearths how inheritance, wealth accumulation, and social mobility reflect the intergenerational struggle to facilitate such progress in the face of systemic oppression.

1. Introduction

The American Dream rests on the idea that the tenets of hard work and effort will satisfy the pursuit of one's desired destiny. For America's Black population, the roots of the American Dream date back to the abolition of slavery, as free Blacks now had, in theory, access to means of social and economic mobility (Baradaran 2019). Yet, freedom did not automatically guarantee free Blacks access to the American Dream or the same opportunities for social advancement that their white counterparts reaped. Additionally, while free Blacks may have had access to those same

opportunities and systems as their white counterparts, the distributions of benefits were unequal—widening such racial divides and promoting toxic socioeconomic gaps. Currently, racial disparities still persist, particularly in financial inheritance rates amongst Black families, the means by which Black families acquire wealth, and how Black Americans struggle to climb the ladder of mobility (Fairlie and Robb 2007). Apprehending how Black families have historically been at an economic disadvantage is necessary to understanding why disparities in wealth accumulation, social mobility, and inheritance currently persist.

This research paper will begin by uncovering how past and present systems advanced and continue to promote racial disparities in social mobility, inheritance, and wealth accumulation. For example, examining housing segregation, racism, crime, and poverty will reveal why Blacks struggle to accumulate wealth. Race-based economic discrimination has deprived Blacks of the right to live in affluent areas—perpetuating the notion of ghettos and limiting their ability to move up the mobility ladder (Baradaran 2019). Additionally, analyzing the failures of Black banks further emphasizes the economic-based discrimination Black Americans faced as they desired to participate in the free market economy. Investigating historical rates in access to schooling will also argue how education, occupations, social networks, and wealth interconnect and how our segregated education system has impacted the ability of Blacks to gain wealth. Grappling with these sociological factors will outline inheritance rates within the Black community and if wealth accumulation is self-made, a product of family inheritance, or both. Lastly, examining previous research pursuits on why Black Americans struggle to acquire wealth will provide an in-depth qualitative and quantitative analysis to support the sociological trends outlined above. Thus, in the pursuit of the American Dream, Black families have experienced generations of racial trauma that has affected their ability to accumulate wealth sufficient to provide stability to future generations.

Subsequently, this research paper will present original, qualitative data to further explore the overarching claims posed beforehand. Data obtained from interviewing seven Black families will result in an analysis of intergenerational wealth distribution and the significance of inheritance, social mobility, and wealth accumulation in those households. Interview questions will range from interpersonal topics related to their childhood, career, education, personal finances, religious beliefs, etc. These interviews will take place over Zoom, and a transcript of subtitles will account for the primary source of data. Obtaining qualitative data will ultimately provide a greater perspective—a firsthand perspective of the intergenerational trauma that Black Americans have faced and continue to face. The questions are in chronological order—meaning one will discuss their childhood and education in the earlier parts of the interview and their adult finances in the latter half. The heart of the interview lies in how each family supports their child in education, housing as an adult, and other pathways that foster generational wealth in contrast to the support they received growing up and, in some cases, the support they continue to receive. Drawing such comparisons will give greater insight into how wealth has been created, maintained, inherited, and, possibly, destroyed throughout three generations of Black families.

2. Historical Analysis of Wealth Accumulation, Social Mobility, and Inheritance

2.1 Wealth Accumulation

Racial wealth gaps and disparities in wealth accumulation have vastly defined the socioeconomic dynamic between Black and white Americans. According to the Pew Research Center, the median white household possessed a net worth 13 times that of the median Black family in 2016—calculated by obtaining the total value of all assets upon subtracting the value of all debts (Pew Research Center 2017). The following year, the U.S. Census Bureau's Survey of Income and Program Participation (SIPP) reported that white households had a median household wealth of $171,700. In comparison, Black families possessed a median household wealth of $9,567. Understanding such overwhelming disparities necessitates examining factors, systems, and conditions perpetuating these gaps. Most notably, the intergenerational cycle of wealth transfer is critical to reproducing such mass wealth inequality, as inheritance patterns transcend generation after generation (Pfeffer and Schoeni 2016). Because of massive discrimination and limitations in gaining wealth, Black communities have historically been unable to accumulate sufficient wealth for future generations (Miller 2011).

The failures of Black banks and Black capitalism reflect how systemic and economic racism barred Black communities as a collective from accumulating wealth. The first Black banks were established after slavery ended during the era of Jim Crow segregation and a climate where racism persisted. They rose to prominence alongside the influx of ghettos in northern cities, pledging to grow and control the Black dollar (Baradaran 2019). Instead of bringing money and wealth into the Black community, Black banks diverted Black deposits out of the community by investing in governmental securities and interests (Baradaran 2019). By financing the mortgages and affairs in other communities, Black banks served as a pipeline that exported local deposit funds to other markets and federal funds. Black banks served as these conduits because they needed to protect themselves from the dangers of lending in the ghetto. They reinvested their customers' incomes in outside communities, as they failed to multiply and grow the dollar in the ghetto. Because of free-market capitalism, Blacks could engage in the capitalist, free-market economy, yet many perceived Black ghettos as an isolated, separate economy—Black capitalism (Baradaran 2019). Black banks had many other liabilities that prevented Black Americans from obtaining substantial wealth. Since many Black Americans deposited small amounts and served economically disadvantaged customers, Black banks spent more money and made less profit through each deposit (Baradaran 2019). Black banking focused on meeting the credit demand for home loans. The collateral for home loans would diminish in value when Blacks purchased those loans, causing the portfolios of Blank banks to suffer (Baradaran 2019). Thus, the liabilities of Black banks and their submission to the inadequacies of the ghetto economy prompted the Black community's failure to accumulate wealth in the twentieth century.

2.2 Inheritance

Inheritance plays a significant role in determining wealth, as those with high incomes in the 1960s were more likely to receive substantial amounts of money (Oliver and Shapiro 2013). Since then, segregation has deprived Blacks of access to education, an expandable network, prosperous career opportunities, and thriving wages—pathways that lead to wealth. Thus, Blacks statistically lacked the backing necessary to obtain high incomes and savings, which barred them from generating and passing down wealth (Chiteji 2010). In a qualitative study published in 1995, Thomas M. Shapiro and Melvin L. Oliver interviewed Black families in Los Angeles to understand the intergenerational transmission of wealth and how inheritance corresponds to financial well-being. The interviews lasted anywhere from forty-five minutes to two and a half hours, and they acquired their interviewees by contacting people, friends, and other acquaintances. They identified a child's formative years, the milestone events of a young adult's life, and the ability to provide childcare for one's children as the primary means inheritance transcends generations (Oliver and Shapiro 2013).

Wealth plays a tremendous role in shaping a child's formative years as education, early friendships, and experiences provide the groundwork for their future financial success, independence, and network quality. Parents use their wealth to enhance their child's cultural capital by nourishing them with high-quality schooling, weeks at summer camps, after-school enrichment activities, and opportunities to participate in sports, vacations, and trips (Harris 2020). Thus, wealthy families give their children access to the same childhood experiences that have shaped them—equipping them with the tools essential to create and maintain wealth later in life (Oliver and Shapiro 2013). The Black families they interviewed noted the advantages of private school education as opposed to public school education because the former exposes their children to families from wealthy economic backgrounds, which allows their children to benefit from a robust cultural capital (Oliver and Shapiro 2013). Additionally, wealth plays a significant role in milestone life events such as marriage, purchasing their first homes, and having children. All the Black families they interviewed depended on financial support from their parents while buying or renting their first home. Going to college—another life milestone—is another influential factor highlighting inheritance disparities, as paying for college tuition instead of borrowing loans accounts for the difference between starting a career with or without the financial toll of education expenses (Oliver and Shapiro 2019). Lastly, Shapiro and Oliver identified a family's ability to provide financial assistance for their grandchildren's childcare as another vital way wealth transcends generations (Oliver and Shapiro 2013). Among the Black families, very few anticipated receiving large sums of inheritance rates—a minimum of $100,000 worth of cash and assets. Those observations assert that—statistically—Black families struggle to provide their children and grandchildren with a robust financial and economic support system.

2.3 Social Mobility

By highlighting how Black families struggled to transmit high intergenerational occupational status, Shapiro and Oliver's research also concluded that race correlates to one’s ability to move upward in the mobility ladder. In their interviews, Shapiro and Oliver compared the various types of occupations—upper-white collar, lower-white collar, upper-blue collar, and lower-blue collar—of parents and their children. They observed that approximately one-third of Black parents from upper-white-collar

backgrounds successfully transmitted their occupation status to their children. Additionally, many of the offspring of parents with upper-white-collar occupations possessed blue-collar jobs. One-half of all Blacks with parents in the upper-blue-collar bracket fell into the lower-blue-collar sector. This failure to uphold occupational position, which prompted many Black families to descend the mobility ladder, represents how Blacks struggle to pass down their social advantage because of their lack of wealth assets (Sykes and Maroto 2016). Black families struggled to ascend the social mobility ladder, as the rates of upward mobility for the offspring of those with lower-blue collar jobs were substantially low (Oliver and Shapiro 2013). Only one-third of Black families they interviewed who grew up in upper-blue collar families moved upward to white-collar occupations. Lastly, less than 30% of Black families born into lower-white-collar families moved upward to upper-white-collar backgrounds.

Ultimately, Shapiro and Oliver's research asserts how Blacks have historically been at a disadvantage in upward mobility and tend to descend on the mobility ladder. But, more importantly, the implications of their research reflect the systems that have prevented Blacks from dominating the occupational and social mobility ladder. Most notably, systemic racism and economic inequality has barred Blacks from having access to affluent schools, occupations, trades, and neighborhoods. Yet, in 2015, Whites with a college degree had $300,000 more wealth than Black families with college degrees, which outlines how equality does not triumph over equity (Oliver and Shapiro 2013). 75% of Black children who grew up in families that lived in poverty remained in the same wealth category as adults (Oliver and Shapiro 2013). Although Blacks legally had the right to receive high wages and white-collar occupations, many do not live in regions with access to such opportunities for economic prosperity and are as poor as they have ever been (Loving, Finke, and Salter 2011). Racial inequality in educational attainment, income, incarceration, and occupational status correlates to the inability of Blacks to gain upward mobility and, ultimately, generational wealth. Black families without extreme wealth often live in underserved communities with limited social networks and inadequate educational resources (McKernan, Ratcliffe, Simms, and Zhang 2014). Many resort to crime in the face of these limited opportunities, which further diminish their chance of social and occupational mobility (Sykes and Maroto 2016). This cycle traps Black families in intergenerational economic hardship and poverty as they lack access to opportunities that promote social mobility (Park, Wiemers, Seltzer 2019). Even those who manage to climb the mobility ladder face the fear of descending or watching their future generations plunge into such cycles of economic hardship because they lack the wealth to maintain their socioeconomic status (Gibson-Davis and Percheski 2018). Thus, the analysis of such research endeavors indicates that the tendency to descend or maintain their footing on the economic mobility ladder is a product of the lack of equity and access to opportunities for acquiring wealth.

3. Methodology

Upon examining how inheritance and mobility disparities have destabilized generations of Black families, it is necessary to understand how Black families seek to inherit their parents' and grandparents' educational, career, and life experiences. It is also vital to acknowledge what types of financial, economic, and social support they aspire to pass down to their children, providing the foundation for wealth accumulation and inheritance. Historically, Black families have struggled to

accumulate and inherit wealth, emphasizing the cruciality of uncovering how Black families overcame these racial and socioeconomic divides and will continue to overcome these burdens in the future. Thus, I obtained qualitative data by interviewing the heads of seven Black families—Mr. Johnson, Mr. Kennedy, Ms. Nancy, Mr. Devin, Mr. Hamer, Ms. Cox, Mr. Jamison—over two weeks to comprehend and enrich these sociological trends. I changed their names for the purpose of anonymity. Like Shapiro and Oliver's interviews, I asked all participants about their experiences grappling with wealth accumulation, inheritance, and social mobility. With each interview taking place over Zoom and ranging from twenty to forty-five minutes, I obtained numerous narratives, stories, and perspectives that unravel how Black families transmit intergenerational support through generations. At the culmination of each interview, I took the transcript of our conversation to compare the various data sets I collected. I obtained my interviewees by asking family members, friends, and acquaintances. Unfortunately, a few families declined to participate in this interview because it required them to speak about personal information. Despite such challenges, seven families expressed interest in being interviewed, which allowed me to collect data smoothly. These Black families were born in the United States, Jamaica, Canada, and several parts of Africa, yet they have been living in the United States, specifically Brooklyn, New York, New Jersey, and the Northeast, for most of their adult lives. The ages of my interviewees ranged from 44 to 73, and they comprise many income and socioeconomic brackets.

Before I began interviewing my participants, I crafted a specialized question list that engaged with what types of financial and intergenerational support Black parents received and how they plan on transmitting similar resources to their children. The official questionnaire is in the Appendix section, which follows the Conclusion. I asked participants about their highest level of education, their parents' highest level of education, the type of financial support they received during their educational pursuits, and the context of their upbringing. Asking these questions allowed me to correlate one's childhood and educational experiences to their ability to accumulate and pass on wealth. After learning about their youth, I pivoted to inquiring about the context of their child's upbringing, their educational expectations for their children, and how they will financially support their children throughout education and beyond. The answers to these responses will enable me to compare the childhood and educational experiences of a parent and their child, which exposed me to how Black families interacted with the mobility ladder. It also revealed how wealth transfers generations, as those born into a family that amassed great wealth raised their children with the same resources and tools for enhanced cultural capital.

The second set of questions asked the participants to speak about the evolution of their career experiences. I asked participants when they obtained their first job and did they get it, how the size of their network changed throughout their careers, and how they plan on supporting their children throughout their careers. These questions sought to discover the relationship between one's network and occupations, as they are indicators of one's wealth. Additionally, these questions aimed to understand how one's childhood and educational experiences, and the context of their upbringing, translate to career aspirations and successes. Bridging these lifetime moments is another indicator of social and occupational mobility, as many managed to find the necessary resources for wealth accumulation during their careers. Lastly, inquiring how one will support their children's career goals will show how Black families will pass on and inherit an expansive network and occupational status—another influencer of social mobility.

The final set of questions inquired about a family's financial background and other life experiences outside their childhood and career that pertain to inheritance, social mobility, and wealth accumulation. I asked participants if they ever received financial support from their parents that they used for their children, if they inherited anything from their parents, and what they anticipate passing down to their children for inheritance. I also asked about the effects of having children on their career and finances, about the role of religion in their lives, and if they had financial help when purchasing/renting their first home. These questions are the heart of the interview because they will specifically answer what inheritance they received and what they seek to pass on. Their answers allowed me to analyze how intergenerational financial, economic, and social support transcends generations in the backdrop of one's childhood, educational, and career experiences. To enhance my understanding of social mobility, I asked my participants if they were willing to share their current family income and estimated family income when growing up as a child. I used five income brackets to classify their differences: (a) $15,000 to $65,000, (b) $65,000 to $100,000, (c) $100,000 to $150,000, (d) $150,000 to $200,000, and (e) $200,000 and above. Measuring the differences in these estimates allowed me to understand how social and occupational mobility intersect with the financial resources one possesses in adulthood instead of childhood. This question is a final attempt to obtain data that might not have appropriately responded to my previous questions about intergenerational support.

Going into these interviews, I wanted more than simple answers to my seemingly simple questions because they evoke greater meaning. I was not hesitant to ask my participants to elaborate or clarify points they made because I wanted to hear their stories, narratives, and voices come to life. I wanted to hear how their unique experiences culminated in their current financial state and their attitudes towards raising and supporting their children with financial inheritance, wealth accumulation, and social mobility.

4. Findings and Qualitative Analysis

4.1 Sociological Trends in Inheritance and Social Mobility

There were many distinct inheritance patterns within the seven Black families I interviewed. Six of the seven interviewees noted that they did not receive any form of financial inheritance; five out of the six reported they plan to pass down financial or asset-based inheritance to their children. I also noticed patterns in inheritance through cultural capital, life milestones, and financial support from grandparents—patterns that Shapiro and Oliver also observed. All seven participants grew up in economically disadvantaged households, as they reported their parents having incomes in bracket A—$15,000 to $65,000. Thus, many did not have access to the cultural advantages that wealthy families possessed. Like many participants, Mr. Johnson emphasized how his parents provided psychological and emotional support, as they lacked the income and wealth to support his education.

"I came from a low income family and they were not able to really support me financially. They did support me intellectually and that pushed me to do the best that I can and go as far as I could."

Ms. Nancy reported a similar narrative by claiming, "um, they were supportive, but, you know, just with, uh, you know, being proud of me and things like that, but that was really it not financially and not helping me with anything." Thus,

emotional, and psychological support made up for the lack of wealth and cultural capital that defines a child's formative years. Ms. Nancy and Mr. Johnson would receive master's degrees, which further emphasizes the impact of emotional and psychological support in the absence of financial support. Four of the seven received a bachelor's degree or higher, while the remaining three only received a High School diploma. Every participant had at least one parent who only received their High School diploma. Yet, every participant agreed that they expect their children to finish their Bachelor's degrees, which reveals that inheritance is not limited to economic and financial support. The families I interviewed transmit intergenerational attitudes about psychological and emotional support for educational success.

Additionally, six of the seven families reported that they did not receive financial support when purchasing or renting their first home—a critical life milestone. All six cited that they saved up for it because they knew their parents were not in the financial position to fund such a cost. Yet, three participants reported that their parents managed to give substantial financial support to their children. For example, Mr. Devin claimed, "only recently my mom offered to pay a portion for my daughter's first year of prep school and we accepted that." Mr. Devin purchased his first home at the age of 23 without the support of his parents, as, at the time, they could not afford to support him. These trends suggest that generation A could not support generation B, but they amassed wealth over time to support generation C—their grandchildren.

Ultimately, these trends display how inheritance impacts a child's educational and formative years, their milestone events, and the grandchildren's early years. Since most of the participants intend on passing on financial inheritance to their future generations, it is reasonable to infer that most climbed the mobility ladder, as many of their parents did not pass down financial inheritance to them.

Mr. Hamer mentioned that he plans for his daughter to inherit "hundreds of thousands of dollars. Um, I still haven't made the money that I have, I have in mind to give my daughter when it's all said, and, but I'm on time. So I give myself to 65 and my goal is to make sure that she has $200,000 in the bank for herself, you know, when it's all said and done for me." He received no financial inheritance from his parents and grew up in an economically disadvantaged household (income bracket A). Thus, his ability to provide substantial financial inheritance for his daughter suggests that he climbed the social and economic mobility ladder.

Patterns in one's ability to move up and down the mobility ladder were also apparent. Every participant reported growing up in a family with an average income in bracket A. Currently, six of the seven families possess an income in brackets C and above, highlighting how they managed to move up the mobility ladder and improve their standard of living. Many reported that they had the opportunity to fund family vacations and recreational activities, which they did not have access to as a child because of their limited resources. Thus, social mobility and cultural capital intersect because as parents move up the mobility ladder, they have access to means of providing their children with the early experiences they never had. Mr. Jamison's ability to finance his children's college education and the childhood experiences he provided them also suggest how social mobility and cultural capital intersect. He noted that "we probably make a little bit too much money in order for the kids to get any financial aid. So we're prepared." He grew up in a family in income bracket A and could not afford the cost of college. Yet, he managed to obtain "a lot of financial aid, um, combined with a scholarship for sports, basketball, and baseball. And then by my junior year, I no longer needed to financial aid because the college or for basketball took off, took over the scholarship." His children are involved in many sports, theater,

and arts, and engaged in many opportunities that bolster cultural capital—further highlighting how social mobility allows families to relieve their children of financial burdens like college.

Mr. Kennedy's story is an example of how downward mobility transcends generations. Mr. Kennedy, who grew up in income bracket A, did not attend college because, while his mother could not afford to support him financially, she "did not actually, um, support my, uh, education, my career goals. She didn't care one way or another, whether I made it or not." Unlike Ms. Nancy or Mr. Johnson, Mr. Kennedy inherited his mother's lack of psychological and emotional support, which contributed to his inability to climb the mobility ladder. Because Mr. Kennedy cannot climb the mobility ladder, he is unable to pass on any financial inheritance to his children. Instead, he aspires to pass on "unpublished writings in the form of a, a book. It explains my five year five careers that I've gone through in my life. And it also explains my way of thinking. Those are the only, that's the only two items I plan on passing down to my, my children." His children will inherit the psychological and emotional support his mother did not give him in his youth.

These results also reflect the research performed by Jasmine Harris, who asserted that wealth transfers to enhanced cultural and social capital. By engaging in experiences such as high-quality schooling and dynamic extracurricular activities, children inherit wealth in the form of cultural capital. Such childhood experiences give them the tools to create and maintain wealth—culminating in mobility—as they grow and develop throughout their adult years. Yet, the inability to transfer such wealth to children does not mean they will not be able to generate wealth. Even though the families I interviewed did not grow up in an environment that emphasized cultural and social capital, they relied on emotional and psychological support to generate wealth, which allowed them to give their children childhood experiences they never had. Thus, emotional, and psychological support is a pathway that provides one generation the means to secure upward social mobility, as they will possess the means to transfer social and cultural capital to the next generation.

4.2 Sociological Trends in Wealth Accumulation

Additionally, there were many wealth accumulation patterns that reflect Shapiro and Oliver's observations. All seven participants noted that expanding their social networks was essential in advancing their careers and building wealth. Mr. Kennedy mentioned how he grew up with "a very large network, very large parish of about maybe seven to 10,000 people in this particular parish. This was a parish of mainly Irish, uh, Italians, Spanish. I was one of the very few black, young men was working as a receptionist for a parish, a Catholic parish. And at the time it was like volunteer work, which actually turned into, uh, another job payment I was paid during this time. Okay. And the network was very, very large. Uh, parish was extremely, very wealthy parish going to church every Sunday. Uh, again, cause I was the receptionist. I would count the Sunday collection on Sundays and that's how I got very much involved with this job as a athletic director." By obtaining his first job through his church, Mr. Kennedy's narrative highlights how growing up with expansive social networks accounts for more job and career options. Similarly, Mr. Jamison expressed, "well, growing up the, the network, obviously when you're a kid and going through college, you don't have as much, but once I got out in the work field, the network, definitely it definitely grew. Um, and then the more stops I did in terms of jobs, each stop, I went the more and more newer people that I met. So, and then that just helped expand, you know, the network." By expanding his network, Mr. Jamison managed to meet others

who helped him advance through his career endeavors. Additionally, Mr. Jamison has an income in bracket E and plans to pass down substantial financial inheritance to his children, which further highlights how expanding social networks correlates to higher income and wealth accumulation.

Because all seven grew up in economically disadvantaged households, many had to build wealth without the assistance of parents. Mr. Hamer's story outlines how he grew up in an environment where he did not learn about the importance of gaining wealth and maintaining individual finances. He stated, "our family never taught us how to save. They didn't teach us about money. Um, because my, even though when I came, my father, my father, mother were together, but they were separate. So my brothers and I we didn't really have that father to, so he didn't talk to me by things like sex and money out, you know, at wealth and all of that. We didn't have those discussions." As he grew older and began his career as a DJ, he noted, "I was able to network with retailers because I was getting product material, um, with all the main label companies. Right. And so that parlay, my radio show became so big that people started selling my tapes tapes of my video show overseas in London. And so when I eventually went back to London, people already knew who I was. Oh wow. And I was able to network that to parlay gigs into Italy." Mr. Hamer started to expand his network as his career took off, allowing him to travel, explore many countries, and expose himself to various cultures. Mr. Hamer has an annual income in bracket C and plans to pass down financial inheritance to his children. Thus, networking leads to more career opportunities, which, in turn, leads to higher income and wealth.

Shapiro and Oliver's research asserted that an expansive network, which many tend to inherit, is essential to building and maintaining wealth. The Black families I interviewed all noted that their networks were influential in advancing their careers and expanding their wealth. The patterns I obtained are consistent with Shapiro and Oliver's analysis of how patterns in social mobility, wealth accumulation, and inheritance intersect.

The results also reflect research performed by Melinda Miller, Ajamu C. Loving, John Salter, and Michael S. Fink. They asserted that discrimination and limitations in gaining wealth barred Black families from accumulating sufficient wealth for future generations. During segregation, Blacks did not have the same opportunities for employment, education, and network accessibility as their white counterparts. Yet, the families I interviewed that displayed upward mobility managed to build a solid network and employment opportunities. Without the limitations of segregation and Jim Crow, as Miller, Loving, Salter, and Fink alluded to, these families had a higher chance of gaining wealth because they had more accessible opportunities.

4.3 Sociological Trends in Religion

The last major trend involves the participant's engagement with religion. Every participant gave a detailed account explaining how religion has played a tremendous role in their life.

Mr. Jamison, for example, claimed, "uh, religion, religion, when I was younger, played a huge role up until probably I was about maybe 30. You say, it's not that it stopped. It just that I feel like once I reached a point where like I was, I was old enough to, you know, kind of not to say, do things on my own, but kind of know how to go about doing things. I think that's when religion stopped, not ne not like a hard stop because even now to this day, I still tell my kids, Hey, listen, you need to have a

relationship, you know, with God... Growing up. I was, you know, my mom was one of those wake up at five o'clock in the morning, six o'clock Sunday morning, wake up, drive two hours to go to church and we'd be in church all day."

Mr. Johnson has a similar testimony when saying, "God is what basically took me to where I am today, as well as all of my kids. Right. Cause my kids, my kids, all of my kids grew up except for my daughter grew up in a low income area, you know, different in, in Brooklyn, New York city. And you know, we did not have a lot of money. We were basically the, for my, most of my kids, we were basically working poor. So God is what allowed them to get to, um, basically to be led to some very good schools from elementary, all the way to high school and college... God has open doors for all of my kids as well as myself and my wife. So definitely. So it's, it's a major, religion is a major part of my life."

Additionally, Mr. Hamer noted that "I played for various, um, organizations in church. Um, right now I play the drums every Sunday in church. And I'm able to, when COVID has happened, I was able to facilitate most of the funerals, um, that go down. I do about three, three to five funerals every week. Wow. Playing the drums and committing the services for the people that have lost loved ones. So I'm part of the ministry that kinda comfort, comforts, the lost ones also, um, I'm in the media ministry because when the pastor preaches my pastors. So what I do is I upload his sermons up to YouTube to make sure that everybody that's not able to attend."

There were many instances when the participants provided a brief narrative about how religion impacts their lives. Because all these families have firm religious beliefs, they possess the attitude and values that compel them to support their children in whatever way possible. Thus, they provide psychological and emotional support, allowing their children to grow up in an atmosphere that champions moral values, which may inspire them to raise their children in a similar manner.

The above analysis of religion reflects the works of Harris, Miller, Loving, Salter, and Fink. In the narratives above, religion is a means of providing an accessible social network, as well as emotional and psychological support. Emotional and psychological support can compensate for the lack of cultural capital. Strong social networks correlate with more opportunities to gain wealth, which Blacks did not possess due to segregation and discrimination. Religion bridges these gaps by providing these families with a strong network and a haven of emotional and psychological support. Children of the recipients of these types of support will also inherit these support systems.

5. Conclusion

The findings of this research paper seek to expand on the concepts of inheritance, social mobility, and wealth accumulation that have historically impacted the Black community's social and economic footing in the United States. Additionally, the qualitative data obtained from these seven interviews expand on the conclusions that Shapiro and Oliver obtained during their research and observations by analyzing psychological, emotional, and financial support, the relevance of social networks in wealth accumulation, and the significance of religion in cultivating intergenerational attitudes. But, more importantly, it showcases how the narratives of these Black families contribute to the desire to live the American Dream. For all the families I interviewed, generation A amassed incomes in bracket A, while most of generation B possessed incomes in brackets C and above. These families have forged through the

challenges of accumulating wealth, moving up on the mobility ladder, and striving to provide an inheritance to ensure that generation C is in a solid economic and financial situation. The data I obtained responds to how each family supports their child in housing as an adult, education, and other means that foster generational wealth in contrast to the financial support they received growing up and, in some cases,—the support they continue to receive.

Besides increasing the sample size of my research population, I seek to vary the demographics of my sample. I need to interview more Black families from various economic backgrounds to bolster my understanding of how disparities in social mobility, wealth accumulation, and inheritance currently affect the Black community. To further the depth of my research, I intend to follow up with the children of the participants I interviewed twenty years from now and interview them to compare their answers to those of their parents. Obtaining such qualitative data will expand my understanding of how social mobility, wealth accumulation, and financial inheritance function within the historical and present context in the lives of Black Americans.

6. Appendix

6.1 Interview Questionnaire

Below is the official list of questions I asked during every interview. Due to privacy and confidentiality concerns, I will not release the official transcripts of these interviews.

6.2 Education

What is your highest level of education and where did you obtain it? (Ask for both partners if applicable)

If the participant(s) obtained a bachelor's degree or higher…

What types of financial support did you receive during your time in college that helped you graduate?

If the participant(s) did NOT obtain a bachelor's degree or higher…

What reasons did you not attend/finish graduating college?

If you were to do it all over again, would you go to/finish college?

Follow up (if needed): Would your parents have supported you financially?

Did your parents support your educational and career goals? How so?

Highest level of education for parents?

What's the highest level of education you expect your children to obtain?

How do you plan on financially supporting your children through college and their educational journey?

… and now we will pivot to more career-specific questions.

6.3 Career

How old were you when you obtained your first job and how did you get it?

Growing up, how big was your "network" and how did that change throughout the course of your career?

Do you have any retirement savings plans? If so, what type? (i.e., 401k, Pension Plans)

How do you plan on supporting your children through their career?

Do you have a checking and savings account?

… and now we will pivot to more family-oriented questions.

6.4 Family

What role, if any, has religion played in your life?

How old were you when you began to live independently? (Away from your parents)

Follow up: What motivated you to live independently?

Did you have financial help did you have when purchasing/renting your first home?

CLARIFICATION- have you ever owned a home?

How old were you when you had your first child?

How did having your child/children impact your career and finances?

Have you ever received financial support from your parents that was used for your children?

Have you received any form of inheritance from your parents?

What do you anticipate passing down to your children for inheritance?

If I were to distinguish between working class, middle class, upper middle class, where would you place your family?

If you're willing to share, what household income bracket would you classify yourself in right now?

A. 15k to 65k

B. 65k to 100k

C. 100k to 150k

D. 150k to 200k

E. 200k and above

Growing up, would you say that letter was the same, lower, or higher?

6.5 Miscellaneous

Parting thoughts: Is there anything you would like to share about how your parents financially supported you and how you plan on financially supporting your children?

References

2017 Data Show Homeowners Nearly 89 Times Wealthier Than Renters. [online] Census.gov. Available at: <https://www.census.gov/library/stories/2020/11/gaps-in-wealth-of-americans-by-household-type-in-2017.html#:~:text=Non-Hispanic%20White%20householders%20had%20a%20median%20household%20wealth,from%20the%20estimate%20for%20non-Hispanic%20White%20householders.%20Education.>.

Baj-Krzyworzeka, M., Szatanek, R., Weglarczyk, K., Baran, J., & Zembala, M. (2007). Tumour-derived microvesicles modulate biological activity of human monocytes. *Immunology Letters*, 113(2), 76–82. https://doi.org/10.1016/j.imlet. 2007. 07.014.

Baradaran, M., 2019. The color of money. Cambridge, Massachusetts: Belknap Press of Harvard University Press.

Chiteji, N., 2010. The Racial Wealth Gap and the Borrower's Dilemma. Journal of Black Studies, 41(2), pp.351-366.

Fairlie, R. and Robb, A., 2007. Why Are Black-Owned Businesses Less Successful than White-Owned Businesses? The Role of Families, Inheritances, and Business Human Capital. Journal of Labor Economics, 25(2), pp.289-323.

Gibson-Davis, C. and Percheski, C., 2018. Children and the Elderly: Wealth Inequality Among America's Dependents. Demography, 55(3), pp.1009-1032.

Harris, J., 2020. Inheriting Educational Capital: Black College Students, Nonbelonging, and Ignored Legacies at Predominantly White Institutions. WSQ: Women's Studies Quarterly, 48(1-2), pp.84-102.

Kochhar, R., Fry, R. and Taylor, P., 2011. Racial Wealth Gaps Rise to Record Highs. From Civil Rights to Economic Justice, 18(2), p.61.

Loving, A., Finke, M. and Salter, J., 2011. Does Home Equity Explain the Black Wealth Gap?. SSRN Electronic Journal,.

McKernan, S., Ratcliffe, C., Simms, M. and Zhang, S., 2014. Do Racial Disparities in Private Transfers Help Explain the Racial Wealth Gap? New Evidence From Longitudinal Data. Demography, 51(3), pp.949-974.

Miller, M., 2011. Land and Racial Wealth Inequality. American Economic Review, 101(3), pp.371-376.

Oliver, M. and Shapiro, T., 2013. Black Wealth/White Wealth: A New Perspective on Racial Inequality. Hoboken: Taylor and Francis.

Oliver, M. and Shapiro, T., 2019. Disrupting the Racial Wealth Gap. Contexts, 18(1), pp.16-21.

Sykes, B. and Maroto, M., 2016. A Wealth of Inequalities: Mass Incarceration, Employment, and Racial Disparities in U.S. Household Wealth, 1996 to 2011. RSF: The Russell Sage Foundation Journal of the Social Sciences, 2(6), p.129.

Pew Research Center's Social & Demographic Trends Project. 2022. Chapter 2: Household Wealth. [online] Available at: <https://www.pewresearch.org/social-trends/2011/07/26/chapter-2-household-wealth/>.

Pfeffer, F. and Schoeni, R., 2016. How Wealth Inequality Shapes Our Future. RSF: The Russell Sage Foundation Journal of the Social Sciences, 2(6), p.2.

Returning Comfort To "Comfort Women": The Effect of Korean Traditional Folk Music on Reactivity and Ethnic Identity

Suhh Yeon Kim

Author Background: *Suhh Yeon Kim grew up in the United States and currently attends Beverly Hills High School in Beverly Hills, California, in the United States. Her Pioneer research concentration was in the field of psychology/culture studies and was titled "Psychology of Immigration."*

Abstract

This mixed methods study examined whether listening to Korean folk traditional music affected heart reactivity and ethnic identity of Korean Japanese immigrant women known as "comfort women" and whether hearing Korean folk traditional music could bring comfort to this traumatized population. The sample consisted of 77 Korean Japanese immigrant women that self-identified as former comfort women. Two different genres of music, Korean folk traditional and classical music, along with no music, were used as the independent variables (IV) to assess the impact on heart rate reactivity and ethnic identity. In the second part of the study, comfort women were interviewed to investigate if interviewees felt that Korean folk music had therapeutic healing abilities. Results from a multivariate analysis of variance (MANOVA) and t-test indicated decreased reactivity and increased ethnic identity after hearing Korean folk traditional music, and interviewed participants reported that listening to Korean folk traditional music gave them a sense of comfort and nostalgia. Results of the study suggest that Korean folk music should be considered for implementation in therapy sessions for traumatized populations such as comfort women.

1. Introduction

Often tricked into going to the Japanese camps with the promise of a well-paid job, Korean military sex slaves, or "comfort women," were the subject of extensive mental and physical abuse. With no formal acknowledgments of wartime atrocities or acceptance of legal responsibilities by the Japanese government, comfort women continue to be a marginalized group as they urge proper reparations whilst struggling with the physical and psychological aftereffects of trauma in manifestations such as posttraumatic stress disorder (PTSD). Music is used in therapy to treat similar victims of trauma diagnosed with PTSD. However, most prior research utilizes Western classical music as a means to evaluate the

therapeutic properties of music, and classical music takes prominence in therapy practice as well. For example, the Bonny Method of Guided Imagery and Music is a conventional therapy method that only uses classical music. However, folk music may present a more personable method of connection to music for members of ethnic groups. In particular, Korean folk traditional music is highly personable in its foundation, as many practices of improvisation, solo singing, and modifications were incorporated into song and performance during colonial times (Pilzer, 2006). Korean folk music continues to serve as a vessel for self-expression and connection to the collective "Korean" identity as it weaves in imagery, historical narratives, and expressions specific and relatable to the culture and people of Korea.

This research is based on the theory that hearing Korean folk music would heighten the curative effects of music for Korean comfort women by increasing ethnic identity levels since music in general is found to have therapeutic effects for traumatized populations. The objective of this research was to find if Korean folk music would be a more effective means to heal and cure Korean victims of past trauma; comfort women were selected as the target participant group as survivors carry a history of utilizing music to cope during their times in captivity. The two dependent variables of the study, reactivity (DV1) and ethnic identity (DV2) were evaluated using a multivariate analysis of variance (MANOVA) and t-test, respectively.

1.1. Brief History of Korean Colonization

From 1910 to 1945, Korea suffered a brutal period of colonization by the Japanese empire. Starting from the invasion of China in 1937 until the end of World War II, nearly 14,000,000 people were murdered under the Japanese military regime, with Koreans and other Asian minorities accounting for an estimated 8,000,000 of the deaths. Furthermore, Japanese imperialist rulings frequently ordered mass burnings of books, historical records, and artwork in an attempt to suppress formations of Korean ethnic and national identities. The postcolonial aftereffects of the Japanese occupation are significant as the Korean culture and population suffered losses in traditions and physical abuse of its people (Yang & Lee, 2016).

1.2. The Sexual Slavery of "Comfort Women"

During the Japanese colonial period, there were about 200,000 Korean women that were forced to work in military brothels as sex slaves, referred to as "comfort women" by the Japanese regime. The Japanese government used highly organized and supervised methods such as abduction and deception to force young women into rape and physical labor. Most comfort women were young girls between the ages of 13 to 19. The women at comfort stations were forced to serve from 10 to more than 50 soldiers a day, and those that resisted were beaten and cut up by guards and Japanese proprietors (Raymond, 2015). Ultimately, an estimated 90% of comfort women died before World War II ended due to the physical abuse and inhospitable living conditions at comfort stations (Blakemore, 2019).

Now, the comfort women survivors continue to suffer from lifelong guilt and stigma from their experiences. The first woman in Korea to report herself as a former comfort woman was Kim Hak-Sun in 1991, nearly 60 years after the comfort station institutions were abolished by General Douglas MacArthur (Wang, 2019). This delay in report was due to the Korean society's shared aversion to discussing the topic of comfort women as well as the perception that comfort women were

voluntary prostitutes during wartime. In consequence, many comfort women felt shame in their experiences, which manifested in various psychological conditions. In particular, studies show a high prevalence of psychiatric disorders such as posttraumatic stress disorder, anxiety, and depression within comfort women and similar survivors of trauma, such as Holocaust survivors (Lee et al., 2019). As there are many women in minority groups that faced similar situations of sexual slavery during World War II, it is important to find better ways to heal for traumatized populations.

1.2.1. Ways of Coping

During colonial times, Korean comfort women utilized music as a means to cope with trauma in detained "comfort stations", or military brothels. Research by Pilzer (2006) recounts how singing traditional folk songs helped maintain the spirits of the comfort women during captive times, reminding them of the nationally shared cultural nostalgia. By using simple musical tunes and instilling complex metaphors into traditional folk song lyrics, the comfort women were able to express their grief and sorrow towards their situation without risking censorship or punishment from the Japanese government. In summary, music was a key method used by different groups of Korean women to preserve ethnic identity, elicit nationalism, and heal during the Japanese colonialist time period.

1.3. The Therapeutic Effects of Music

Korean immigrant women have also historically performed Korean folk music as a form of therapy to maintain ethnic enclaves and cultural communities. Choi (2021) details that the Korean diaspora in Hawaii during the Japanese colonial period formed groups to stage performances of Korean folk music, songs, and dances, displaying their cultural songs and costumes for tourists while seeking to preserve their cultural legacies as immigrants in a foreign country.

Korean folk music was one tradition that could not be tangibly destroyed during Japanese colonial times. Thus, many independence movement activists and vulnerable citizens that participated in the Korean diaspora as a result of Japanese colonialism depended upon Korean folk songs to preserve their ethnic identity and bolster nationalistic sentiments by altering lyrics and tunes to fit the narrative during difficult times. Despite the significance of Korean folk songs in continuing the Korean independence movement during colonial times, there is minimal research regarding their connection to ethnic identity or their ability to elicit feelings of patriotism and connection to Korea.

Research supports the use of music as an effective healing source in therapy sessions. In a study with adult victims of PTSD, it was reported that individuals reported feeling calmer and at ease after music therapy sessions (Landis-Shack et al., 2017). Studies also found decreases in individual stress levels as a result of changes in hormones and mood after listening to music. There are also many existing music therapy methods such as the Bonny Method of Guided Imagery and Music, supporting that music is a notable method of healing victims of trauma and psychological abuse.

1.3.1. Music and Ethnic Identity

Research further supports that music, especially traditional forms such as folk music, are closely linked to ethnic identity developments within minority groups (Good et al., 2020). Hearing folk traditional music and immersing oneself with a

connective tie to tradition was found to heighten feelings of belonging with individual ethnic groups. Increased ethnic identity and national attachment is also positively linked to wellbeing, especially among Asian populations (Zdrenka et al., 2015). Therefore, folk traditional music holds potential in presenting therapeutic effects to Korean populations and comfort women as it could heighten ethnic identity levels to enhance wellbeing.

1.4. Heart Rate Reactivity

Heart rate is a common indicator for reactivity. Research indicates that heart rate is a good indicator for stress levels within individuals (Thayer et al., 2011). Quantitative studies suggest that higher levels of heart rate are a good measure for both acute and long term stressors, especially among victims of trauma and PTSD (Shalev et al., 1998).

1.5. Goals of the Current Study

There were two main goals of this study. First, I assessed if there was a relationship between ethnic identity and folk music with Korean comfort women living in Japan, and then if listening to Korean traditional folk songs can produce greater curative effects for traumatized populations such as comfort women. This mixed methods study examined this by first measuring the ethnic identity level of participants while there is no music playing in the background, and comparing these results to ethnic identity levels measured after participants listened to Korean folk traditional music containing messages of nationalism and the Korean independence movement.

Hypothesis 1. Listening to Korean folk traditional music will increase ethnic identity for participants.

Hypothesis 2. A lower heart reactivity will be measured upon listening to Korean traditional folk songs as opposed to hearing classical music or no music.

Hypothesis 3. Participants in the focus group will report that Korean folk traditional music generates a greater sense of comfort compared to classical music and possesses curative effects for traumatized populations.

2. Method

2.1. Participants

Participants for the study were Korean former comfort women living in Japan. The study had a sample size of 77 participants between the ages of 82 and 99, born during the Japanese colonization of Korea (1910-1945). The average age of the recruited participants was 89.2 (+ SD = 3.47). The drop out rate of the study was 7%, and the refusal rate was 2%. In order to minimize the dropout rate, I contacted people that were familiar with potential candidates of the research.

2.1.1. Demographics

Participants lived an average of 76.2 yrs in Japan (+ SD = 3.21) and served as “comfort women” for an average of 5 years (+ SD = 2.52). On average, participants were 17.8 years old during time served as comfort women (+ SD = 3.47).

2.2. Materials/Measures

All materials were administered in Korean as this was the main language for the participants. If measures were not previously available in Korean, measures were back-translated and normed prior to administration. One of the research assistants, a graduate student with supreme fluency in Korean and English, was instructed to first translate all scales and measures into Korean. Then, the other research assistant back-translated the scales and measures into English. The two versions of the scales and measures, one originally in English and one back-translated, were compared to assess accuracy.

2.2.1. Music

The Korean folk traditional songs that were used in this study are "*New Arirang*," "*Ilpyeon Danshim*," and "*Ae Guk Ga*". The three classical pieces that were used in this study are "Für Elise" by Beethoven, "The Blue Danube" by Johann Strauss, and "Air on the G String" by Bach. These songs were chosen based on the results of a preliminary focus group (discussed below).

2.2.2. Reactivity

Heart rate was used as a measure for reactivity and was measured in beats per minute (bpm). At the beginning of the study, participants were given a wireless heart rate monitor manufactured by Meditech that was secured around the wrist. The digital heart rate trackers consistently recorded participant heart rates every five minutes from the beginning to the end of the study. In this experiment, higher differences in heart rates before and after listening to music indicated higher reactivity, while lower differences indicated lower reactivity. Higher reactivity or faster heart rates are related to increased levels of stress and restlessness as evidenced in previous research (Shale et al., 1998). Meanwhile, a lower heart rate is connected to feelings of stability and calmness amongst individuals (Oneda et al., 2010). Therefore, heart rate would be a valid and consistent measure of the effects that music could have on calming and lowering the stress levels of the participants in this study.

2.2.3. Ethnic Identity

An abbreviated version of the Multigroup Ethnic Identity Measure (MEIM) by Phinney (1992) was used to measure ethnic identity. In particular, the Affirmation and Belonging subscale was used to measure ethnic pride and the level of belonging and attachment felt towards one's ethnic group. This subscale includes questions such as "*I have a strong sense of belonging to my own ethnic group*," and "*I feel a strong attachment towards my own ethnic group*". The affirmation scale's items were rated on a scale from 1, *strongly disagree*, to 4, *strongly agree*, with higher scores representing a greater sense of ethnic identity felt by the participants and a lower score indicating a lower sense of ethnic identity. The MEIM was chosen because it is a general ethnic identity measure that has been used before with multiple groups. However, in order to make it more comprehensible to the current sample, certain phrases were modified. The MEIM indicates good internal consistency, with a Cronbach's alpha of 0.89 for the Affirmation and Belonging subscale (Phinney and Ong 2007). The scale is valid as lower scores are correlated with longer periods of time lived in Japan or higher levels of acculturation, $r(78) = 0.45$, $p < .01$.

2.3. Procedure

This study was approved by The Central Institutional Review Board of South Korea and participants provided written consent in Korean. All of the 77 recruited participants were registered former comfort women survivors who settled in Japan after the Korean liberation. Participants were recruited from various senior centers across Japan, especially ones that catered towards older Korean populations. Flyers, which included an overview of the study in both Japanese and Korean, were distributed at partnered senior centers. Additionally, in-person visits and emails were sent out by senior centers with the same information along with the author's name, number, and email for questions or concerns about the research. For data collection, each participant was assigned a number, and data for each participant was stored on a computer that was password protected.

Participants were eligible if they met the following characteristics: (1) they self-identified as Korean Japanese that previously served as comfort women (2) they had no hearing difficulties or extenuating health concerns. To assess this, participants were required to take a hearing test and confirm that they had no special heart conditions, in order to ensure study results would be unaffected by extenuating circumstances (e.g. loss of hearing, high blood pressure). Participants were given a paper consent form outlining the basic procedures of the research and informed that they would partake in research that compared memory and recall abilities to heart rate reactivity. After the study, participants were debriefed about the original purpose of the study.

Two lab assistants fluent in Korean were employed throughout the study. Lab assistants were uninformed about the purpose of the study and were instructed to guide participants in sitting down, putting heart rate trackers around their wrists, and distributing pencils for paper assessments and surveys. The entire study lasted for 50 minutes with time periods divided into five 10 minute blocks. To help with participation, participants were provided with transportation to the research site, costs for transportation were covered, and all participants received a 4,000 yen voucher for groceries as compensation. Additionally, all participants were connected to free resources for therapy and Korean programs for music making and learning new instruments.

2.3.1. (IV) Manipulation Check

In order to assess the effectiveness of the independent variable, a focus group was conducted prior to the research with a separate group of similar participants. In this group, participants were asked to nominate three Korean traditional folk songs that they believed would invoke nationalistic sentiments and memories of their native country. The initial list of traditional folk songs consisted of 20 songs that carried altered lyrics as a result of the Korean independence movement. The nominations were based on 1) familiarity with the musical tune, and 2) familiarity with the lyrics of the Korean folk song. The songs chosen were: "*New Arirang*," "*Ilpyeon Danshim*," and "*Ae Guk Ga*." For classical music, three songs were chosen at random from a list of 300 classical music compositions through computerized generations. The chosen pieces were: "Für Elise" by Beethoven, "The Blue Danube" by Johann Strauss, and "Air on the G String" by Bach.

2.3.2. Experiment

Prior to beginning the experiment, all participants were fitted with a wireless heart rate monitor on their wrists that measured heart rate throughout the various blocks

of the experiment. This experiment had five blocks with 10 minutes each, and the entire study lasted for 50 minutes. In block one, participants filled out the Affirmation and Belonging subscale from the MEIM with no music playing in the background. In block two, an elementary level puzzle test was administered, during which classical music played in the background. For block three, participants were given a 10 minute break with classical music playing in the background. Following this, Korean folk traditional music was played in the background during blocks four and five. During block four, participants were administered a second puzzle test. In block five, participants once again filled out the Affirmation and Belonging subscale from the MEIM after being exposed to Korean folk traditional music for twenty minutes. All measures were taken through pencil and paper. The filler puzzle tests were not correlated with the results of the ethnic identity scale or heart rate reactivity $r(78) = 0$, $p < .01$.

2.3.3. Debriefing

After the experimental component of the research, participants were informed about the purpose of the study. It was explained that heart rate was measured throughout the experiment to assess reactivity upon hearing Korean traditional music. The purpose of comparing classical music, the conventional choice for music therapy, with the effects of Korean traditional folk songs was explained. Participants that were a part of the post-experiment focus group were debriefed after the interview session.

2.3.4. Post-experiment Interview

A focus group lasting for 60 minutes was conducted after the research with 12 comfort women survivors to collect qualitative data on the therapeutic effects of Korean traditional music compared to classical music or no music. The 12 participants were chosen through the random generator method; there were no dropouts or refusals. Participants were asked if they felt a greater sense of comfort upon hearing Korean traditional folk music compared to classical music. Beliefs about therapy treatments with Korean traditional folk music were also collected. Participants were also asked to describe their experiences with Korean traditional folk music and their encounters with it during colonial and postcolonial time periods. Eight questions pertaining to these topics were originally written in English then back-translated to Korean to assess accuracy (See Appendix). The interview was recorded on an audio recording device and conducted in Korean.

3. Results

This was a mixed methods study that assessed the effect of music on reactivity, as indexed by heart rate, and ethnic identity. For the quantitative portion of study, a one-way MANOVA and a t-test were used, whereas an interview was conducted for the qualitative component of the research. The effect of music choice was examined on two outcome variables, reactivity (DV1) and ethnic identity (DV2).

3.1. Descriptive Results

The study found that there was a strong positive correlation between lower levels of reactivity to Korean folk traditional music and higher levels of ethnic identity, r(55)

= .49, p < .01. Fewer years of residency in Japan correlated with greater amounts of increased ethnic identity after hearing Korean folk music r(77) = 0.44, p < .01.

3.2. Inferential Results

When participants were exposed to Korean music, they had significantly lower reactivity, or heart rates, than when exposed to other types of music (See Figure 1). At rest, when participants heard no music, the average heart rate was measured to be 101 bpm (+ SD = 1.28). When classical music was played in the background, participants had a slightly lower heart rate of 97 beats per minute (+ SD = 3.44). Once Korean traditional music was played, participants had the lowest average heart rate of 70 beats per minute (+ SD = 3.47). Data after the first five minutes of the experiment when participants engaged in a puzzle activity was not used, as arousal was expected. The heart rate sum score of the intervals prior to intervention and after intervention was added then averaged out. In both scenarios, participants had a lower heart rate upon hearing music compared to when they were exposed to no music. The differences in heart rate supports the hypothesis that hearing Korean folk traditional music would lead to a decreased level of reactivity from participants. Furthermore, for any given score, Korean traditional music tended to have lower levels of reactivity than classical music. Figure 2 graphs the heart rate for one participant in the research. There is a sharp decrease after minute 20, which is when Korean folk traditional music began to play. The MANOVA values for this suggest significant effects of Korean folk music on reactivity, F(2,74)=VALUE, p< 0.05 eta square. A post hoc Tukey test shows that heart rate decreased most significantly upon listening to Korean folk traditional music at p < .05.

The responses on the Phinney's ethnic identity subscale of Affirmation and Belonging indicated that participants felt an increased level of connection to their ethnic group after being exposed to Korean music compared to not listening to Korean music (See Figure 3). Prior to exposure, participants had an average score of 2.9 out of 4 on the Phinney's subscale of Affirmation and Belonging, indicating moderate levels of ethnic identity. After listening to Korean folk music, participants scored a higher average score of 3.7, indicating a notable increase in the level of ethnic identity and feelings of belonging within participants (+ SD = 2.57). This supports the hypothesis that Korean folk music would influence feelings of belonging with one's ethnic group. The t-test values for this suggest significant effects of Korean folk music on ethnic identity, t(47) = 2.4, p< 0.05. A post hoc Tukey test shows that ethnic identity increased significantly upon listening to Korean folk traditional music at p < .05.

3.3. Qualitative Analyses - Positionality and Validity

Following the quantitative analyses, a qualitative study was conducted. In all research, it is necessary to understand our positionality, or perspective on the gathered data. For this study, it is important to note that the author is a second generation Korean American with Korean Japanese immigrant grandparents. Her mother is a pianist who frequently played Korean folk traditional music throughout her childhood, and the author is a violinist who performs reinterpreted Korean folk traditional music. Consequently, Korean traditional folk music is a prominent factor that contributed to developing a connection to the author's cultural identity. As the author's musical and ethnic background may have influenced her interpretations of the data, she made sure to share her interpretations of the study with the

participants in order to increase its validity.

3.4. Focus Group

First, a focus group was conducted with a random selection of the participants who served as comfort women during Japanese colonial times to gather insight into the role of Korean traditional music in the former comfort women's lives and to assess whether Korean folk songs carried therapeutic effects for traumatized participants. A thematic content analysis was used to analyze data from the transcribed interviews. The process of familiarization, naming of categories, and interpretation of themes was used. The 12 participants who identified as comfort women discussed the common theme of Korean traditional folk music being used as a means to relieve trauma and strengthen resilience during colonial times.

Next, after conducting the focus group, a transcript of the focus groups was obtained and a thematic analysis coding was run in order to find recurring themes and patterns between participant responses.

3.4.1. Descriptive Results

The identified themes centered on the role of Korean folk music in maintaining ethnic identity and nationalist spirit during colonial times. Out of 12 interviewed participants, 10 responded that Korean traditional music played a significant role during their time serving as comfort women. Over half (67%) of the participants agreed that Korean traditional folk songs helped strengthen ties to their ethnic identity.

All of the participants responded that they felt a greater sense of comfort when hearing Korean traditional music compared with other forms of music. One participant related that "hearing the lyrics help [her] remember and honor the difficult times." Participants agreed that the unique characteristics such as rhythm, beat, timbre, and lyrics of Korean traditional folk songs help differentiate them from other types of music to deliver a greater sense of comfort. Eight participants mentioned that traditional folk tunes allowed for the continued battle of comfort women to win freedom and bolstered nationalist sentiments as well as ethnic identity during a difficult time of captivity. Three participants reported that singing Korean songs possessed healing components during colonial times because it allowed them to cope with their stories and painful memories in ways that other genres of music cannot convey.

When themes began to be repeated and participants felt that the topic was sufficiently covered, saturation was achieved.

4. Discussion

The purpose of this study was to assess whether listening to Korean folk music triggers reactivity within Korean Japanese immigrant comfort women and if hearing Korean traditional music increases ethnic identity. The results of the study support the hypotheses and indicated that 1) Korean Japanese immigrant women experienced lower reactivity upon hearing Korean folk traditional music compared to hearing classical music or no music, 2) ethnic identity and feelings of belonging increased after listening to Korean folk traditional music, and 3) participants reported that Korean folk traditional music generated a greater sense of comfort

compared to classical music and possesses curative effects for traumatized populations.

The research suggests that exposure to Korean music promotes lower reactivity than hearing classical music or no music. Heart rate in beats per minute decreased the greatest amount for all participants upon listening to Korean folk traditional music, suggesting that participants were in their calmest and least stressed physical states after hearing Korean folk traditional music. The heart rates for all participants after hearing both Korean traditional and classical music was lower than when hearing no music, which further supports the therapeutic abilities of music to treat PTSD and anxiety as it is able to calm one's heart rate (Khanade and Sasangohar 2017). Although the exact causal relationship cannot be concluded from this research, this phenomena could be the result of consolation from hearing patriotic lyrics from Korean folk music that touch upon the themes of preserving a Korean identity, the Korean independence movement during colonial times, and descriptions of the Korean homeland. Participants would be reminded of their childhood experiences in Korea prior to coming to Japan, and lower reactivity could be stirred from nationalistic sentiments or feelings of longing and nostalgia about Korea. Previous research supports this conclusion as quantitative studies have indicated that comfort women felt the highest satisfaction with life prior to being forced into servitude or being abducted to Japan (Lee et al., 2017). It is also noted that the comfort women regard their childhood in Korea with cherished sentiments and feel deeply consoled when discussing their lives back in Korea with family prior to confinement. The traditional melodies of Korean folk music could also affect reactivity as participants could resonate with the timbre of traditional Korean instruments or the rhythms and tonal patterns heard throughout the songs.

Comparing results from the Phinney's ethnic identity subscale of Belonging and Affirmation before and after listening to Korean folk music indicates that 74 out of 77 participants experienced an increase in their level of belonging felt with their Korean ethnic identity. This suggests that listening to Korean folk traditional music may be one factor that can boost levels of ethnic identity and feelings of belonging with one's Korean heritage. The folk songs that incorporate Korean lyrics and messages about promoting Korean solidarity amongst difficult times may have influenced this outcome. Participants could be reminded of their Korean connections through hearing the Korean folk music, eliciting memories of when they believed life to be most satisfactory. As demonstrated by prior research, increased ethnic identity results in heightened wellbeing, especially within Asian populations (Zdrenka et al., 2015). This suggests that Korean folk music can be a crucial component to increase the wellbeing of traumatized populations such as comfort women as it increases levels of belonging with ethnic identity and culture.

The qualitative research data suggest that participants felt a greater sense of comfort upon hearing Korean folk traditional music compared to the classical music that is most conventionally used in music therapy. As Korean folk songs traditionally allow for flexibility and variability in improvisations, comfort women are able to weave in ambiguous metaphors and sentiments within their song to convey their true emotions while maintaining attachment to their Korean cultural identity. Especially in the post-colonial era, in which the history of sexual enslavement of Korean women was often shamed by society, traditional folk tunes most likely allow silenced comfort women who were silenced to honor their sacrifices and remember the pains of the past. Korean traditional folk music both heals and bolsters a sense of ethnic and cultural identity within comfort women.

Interviewees also reported that Korean folk traditional music instilled

dignity and honor back into their painful memories and self-perceptions. The Korean society overall held a sense of shame and aversion to the topic of military slavery and comfort women until the late 1980s (Pilzer, 2006). In consequence, many former comfort women were led to believe that they had been willful prostitutes serving the Japanese government and held the memories often labeled as "shameful" in secret to fit into society and their families once returning. The interview and research suggests that Korean music may activate several psychological reactions in this stigmatized group, aiding to restore honor and self-respect rather than feeling ashamed of their experiences in sexual slavery.

This study is significant because it adds to the body of research on women who have been sexually subjucated and provides clues into how we can provide them care. It also adds to the understudied research on Korean folk music and its connection to the preservation of ethnic identity. There is little quantitative research conducted with comfort women, especially research that pertains to Korean folk music and its curative effects for traumatized populations. It is significant that Korean folk music may present a better method in the healing process for comfort women and vulnerable Korean populations.

4.1. Limitations of the Study

As the population of Korean "comfort women" is very small with only 151 registered survivors currently in Korea, this study is most likely representative (Lee 2018). However, the study was based on women who chose to participate in this study. As such, it was not a randomized sample, which may limit its generalizability. Additionally, the drop out rate was 7%, meaning that a small but significant number of potential participants did not wish to join the study, which may have affected the results. Despite this limitation, however, those who were exposed to music in the study showed uniform increases to ethnic identity and decreased heart rate reactivity to indicate good consistency in results.

Another point to consider is that some participants may be naturally more empathetic, meaning that their reactions to hearing the Korean folk music may not necessarily be a result of a sense of connection to the Korean identity. In future studies, individual variabilities in empathy could be measured in order to account for intrinsic levels of empathy within participants that could influence results. However, although there was variability within individuals, this research uniformly found that hearing Korean music results in higher reactivity and increased levels of ethnic identity than hearing classical or no music.

4.2. Future Recommendations

The integration of Korean folk traditional music into therapy should be considered to diversify ways to provide relief in traumatized populations, especially ethnic minority groups that were culturally oppressed. Additionally, the effects of actively producing Korean folk traditional music on ethnic identity, reactivity, and its healing properties can be investigated in future research. The activation of music often promotes solidarity amongst immigrant populations as evidenced by the Korean diaspora in Hawaii, where Korean women formed groups to perform Korean folk music and preserve connections to ethnic heritage (Choi, 2021). The effect of actively engaging in creating traditional music as well as building a community to share the music could be a cure for trauma such as that experienced by comfort women and other war survivors.

Future research on this topic is recommended to incorporate a greater variety of Korean folk music utilized during the independence movement into the experimental design. Additionally, levels of acculturation can also be examined. With increased levels of acculturation to the host society, in this case Japan, the Korean Japanese immigrants may feel a diminished sense of sympathy and connection when listening to the Korean traditional folk music. The sample can also be expanded beyond comfort women to examine Korean Japanese immigrants and the ability of Korean folk music to boost connection to one's ethnic heritage to promote wellbeing.

4.3. Policy Implications

The healing properties of Korean folk music could be used to provide support for comfort women who were traumatized as a result of Japanese colonial rule. The interviewed women indicated that they felt a sense of connection to their ethnic heritage upon listening to the Korean folk traditional music. Stronger connection to ethnic identity is strongly linked to physical and mental wellbeing; therefore, this study provides evidence of Korean traditional music as a potential tool for speeding the process of healing both traumatized and vulnerable Korean populations.

Today, the Japanese government still has not properly delivered justice to comfort women, despite 70 years passing since the end of World War II. No formal acknowledgments have been made regarding the human rights law violations committed by Japan, and remaining survivors continue to advocate for an adequate apology from the Japanese government. As a means to compensate and assume responsibility for war crimes, the government of Japan can provide support for comfort women and fund music therapy programs that incorporate elements of Korean folk traditional music.

Furthermore, the results of this study emphasize the importance of the continued study of Korean folk traditional music. In April 2022, the Ministry of Education in Korea altered the curriculum for music education in elementary, middle, and high schools, and the new curriculum lacked any guidelines previously present for teaching Korean folk traditional music. Although there were numerous protests from gugak artists, or Korean folk traditional music performers, the new curriculum is prepared to come into effect in 2025, and schools will no longer be required to teach Korean folk music. Greater funding and a change in policy must be implemented to preserve Korean folk music as it possesses significant curative effects and aids in increasing levels of Korean ethnic identity and feelings of belonging, all of which contribute to boost wellbeing within individuals. In sum, Korean traditional music can be a valuable component in healing the past traumas of comfort women subject to sexual enslavement during colonial times and should be promoted both by governments and through integration in music therapy programs.

References

Atkins, E. T. 2007. "The Dual Career of 'Arirang': The Korean Resistance Anthem That Became a Japanese Pop Hit." The Journal of Asian Studies 66 (3): 645–87. https://doi.org/10.1017/s0021911807000927.

Blakemore, E. 2019. "The Brutal History of Japan's 'Comfort Women.'" HISTORY. 2022. https://www.history.com/news/comfort-women-japan-military-

brothels-korea#:%7E:text=Records%20of%20the%20women%27s%20subjugation.

Choi, H. (2021). Curating koreanness: Musical activities of elite korean women in Hawai'i during the Japanese colonial period, 1910–1945. *Women & Music*, 25, 110-127.

Good, A., Sims, L., Clarke, K., & Russo, F. A. (2020). Indigenous youth reconnect with cultural identity: The evaluation of a community- and school-based traditional music program. *Journal of Community Psychology*, 49(2), 588–604. https://doi.org/10.1002/jcop.22481

Howard, K. (1999). Minyo in Korea: Songs of the people and songs for the people. *Asian Music*, 30(2), 1–37.

Yang, J., & Lee, S.-H. (2016). Arirang: How did the folk music promote solidarity during a period of colonization and diaspora. *Serials Journals*.

Kim, S. S. H. C. (2017). Korean "Han" and the postcolonial afterlives of "The Beauty of Sorrow." *Korean Studies*, 41, 253–279.

Koo, S. (2019). Zainichi korean identity and performing North Korean music in Japan. *Korean Studies*, 43, 169-195.

Landis-Shack, N., Heinz, A. J., & Bonn-Miller, M. O. (2017). Music Therapy for Posttraumatic Stress in Adults: A Theoretical Review. *Psychomusicology*, 27(4), 334–342. https://doi.org/10.1037/pmu0000192

Lee, J., Kwak, Y. S., Kim, Y. J., Kim, E. J., Park, E. J., Shin, Y., Lee, B. H., Lee, S. H., Jung, H. Y., Lee, I., Hwang, J. I., Kim, D., & Lee, S. I. (2018). Psychiatric Sequelae of Former "Comfort Women," Survivors of the Japanese Military Sexual Slavery during World War II. Psychiatry investigation, 15(4), 336–343. https://doi.org/10.30773/pi.2017.11.08.2

Lee, J., Kwak, Y. S., Kim, Y. J., Kim, E. J., Park, E. J., Shin, Y., Lee, B. H., Lee, S. H., Jung, H. Y., Lee, I., Hwang, J. I., Kim, D., & Lee, S. I. (2019). Transgenerational Transmission of Trauma: Psychiatric Evaluation of Offspring of Former "Comfort Women," Survivors of the Japanese Military Sexual Slavery during World War II. Psychiatry investigation, 16(3), 249–253. https://doi.org/10.30773/pi.2019.01.21

Oneda, B., Ortega, K. C., Gusmão, J. L., Araújo, T. G., & Mion, D. (2010). Sympathetic nerve activity is decreased during device-guided slow breathing. Hypertension Research, 33(7), 708–712. https://doi.org/10.1038/hr.2010.74

Pilzer, J. D. (2006). "My heart, the number one": Singing in the lives of South Korean survivors of Japanese military sexual slavery (Order No. 3231443).

Raymond, J. G. (2015). Honoring the "Comfort Women" Drafted into Military Sexual Slavery. Radical Feminist Conference.

Thayer, J. F., Åhs, F., Fredrikson, M., Sollers III, J. J., Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: Implications for heart rate variability as a marker of stress and health. Neuroscience and Biobehavioral Reviews, 36(2), 747-756. doi:10.1016/J.NEUBIOREV.2011.11.009

Shalev, A. Y., Sahar, T., Freedman, S., Peri, T., Glick, N., Brandes, D., Orr, S. P., & Pitman, R. K. (1998). A Prospective Study of Heart Rate Response Following Trauma and the Subsequent Development of Posttraumatic Stress Disorder. Archives of General Psychiatry, 55(6), 553. https://doi.org/10.1001/archpsyc.55.6.553

Wang, Q. E. (2019). The study of "Comfort women": Revealing a hidden past—introduction. Chinese Studies in History, 53(1), 1–5. https://doi.org/

10.1080/00094633.2019.1691414
Zdrenka, M., Yogeeswaran, K., Stronge, S., & Sibley, C. G. (2015). Ethnic and national attachment as predictors of wellbeing among New Zealand Europeans, Māori, Asians, and Pacific Nations peoples. *International Journal of Intercultural Relations*, 49, 114–120. https://doi.org/10.1016/j.ijintrel.2015.07.003

Figures and Tables

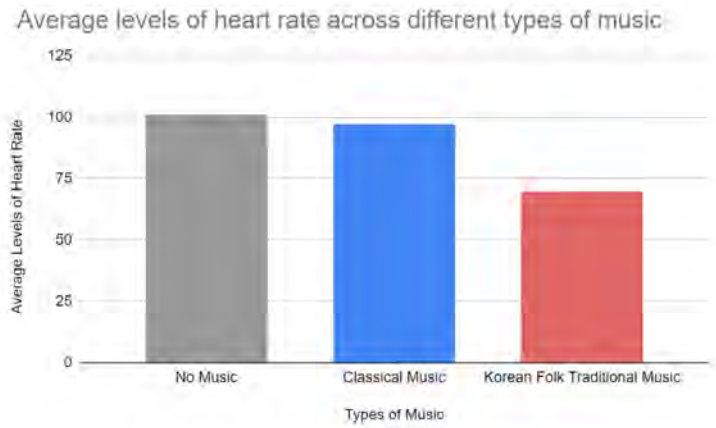


Figure 1. *Average levels of heart rate across different types of music.*

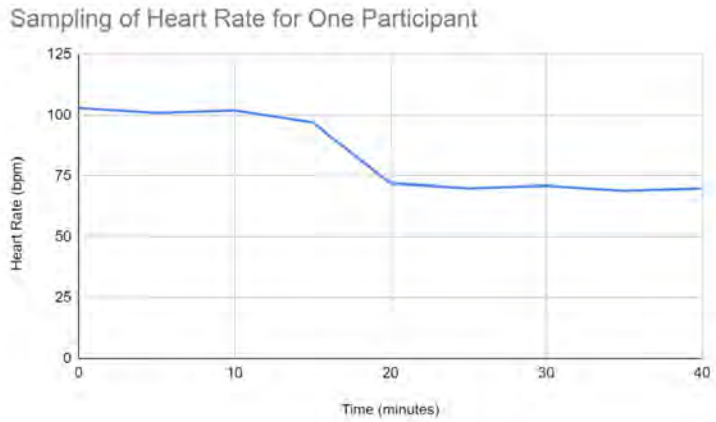


Figure 2. *Sampling of heart rate for one participant*

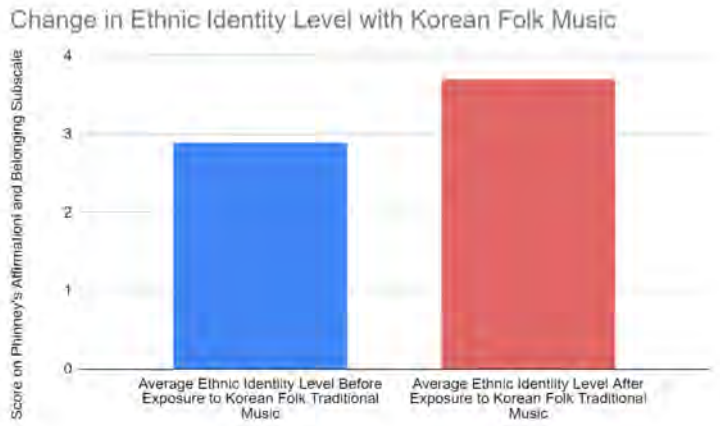


Figure 3. *Change in ethnic identity level with Korean folk music*

Table 1.

Themes	Definition of theme	Quote
Reminder of Korea as homeland	Participants indicate that listening to Korean folk music reminds them of their life back in Korea prior to moving to Japan to serve as comfort women.	"[The music] is the sound of my country, my people. I resonate with it. It's like hearing the mountains and landscapes in the timbre and melodies"
Nostalgia for childhood	Many participants note that they feel at peace thinking about their family and siblings but also nostalgia for lost members.	"Hearing it [Korean folk music] reminds me of the good times, when my mother used to sing with me."
Honor	Korean folk traditional music reinstills dignity and honor back to comfort women as brave women that sacrificed themselves for their country.	"Hearing the lyrics helps [me] remember and honor the difficult times instead of feeling ashamed."
Rhythm and timbre of Korean folk music	Participants reported that they prefer the rhythm and timber of Korean folk music as it plays on traditional methods of sound making that they are familiar with.	"It's just more comforting to hear; I know the different instruments that make up each of the sounds."

Themes	Definition of theme	Quote
Connection to lyrics	For many of the comfort women, the lyrics of Korean folk music carry messages and personal anecdotes that they could relate to.	"The lyrics are what most touched me, and help me remember and cope through my painful memories."
Enhanced Wellbeing	Participants reported that Korean traditional music helps them feel more relaxed and at ease compared to just hearing classical music.	"Classical music is good but listening to Korean music is better and helps me feel more relaxed. I can relate to it better."

Appendix

The questions were back-translated by research assistants to assess accuracy and asked in Korean during the qualitative research.

Questions

1. How familiar are you with Korean folk music?
2. Do you have any experience with it in your childhood?
3. Did music play any part during your time coping with your experiences in Japan?
4. How did hearing Korean folk music make you feel?
5. Do you enjoy listening to Korean traditional music?
6. How connected do you feel to Korea?
7. What role, if any, does Korean music play in your life?
8. If you had to explain to someone how you feel when you hear Korean music, what would you say?