volume 7 | spring 2024

UNC JOURNEY

The Journal of Undergraduate Research at the University of North Carolina Chapel Hill

Letter From the Editors

Dear JOURney Readers,

Welcome to this year's edition of JOURney, the University of North Carolina at Chapel Hill's premier interdisciplinary journal, showcasing the vibrant undergraduate research across our campus.

Since its inception by our founder, Gabi Stein, JOURney has aimed to illuminate the scholarly pursuits of UNC-Chapel Hill undergraduates. Supported by the Office of Undergraduate Research, our publication has become a vital platform where students can share their SURF projects, portions of their honors theses, and independent studies. This year has been particularly inspiring, as we received a record number of submissions, reflecting our student body's dedication and intellectual curiosity.

We are thrilled to present 13 original articles in this volume that delve into pressing, intriguing, and complex research topics. Each piece is a testament to its author's rigorous work and passion, and we extend our deepest congratulations to all the students who contributed.

Our gratitude extends to everyone who has played a role in sustaining JOURney's mission to foster and celebrate the early-stage research endeavors of our students. Thank you to the Office for Undergraduate Research for its steadfast support. We also immensely appreciate our diligent editorial board and publicity team, whose tireless efforts have made this edition possible. As your Co-Editors-in-Chief, we have been privileged to witness and contribute to the ongoing growth of this exceptional organization.

We invite you now to explore the diverse and insightful research featured in this issue. We hope it sparks curiosity, inspires further inquiry, and enhances your appreciation for the academic talents cultivated here at UNC-Chapel Hill.

Warmest regards,

Ricardo Tieghi and Roshni Arun
Co-Editors-in-Chief, JOURney
2023-2024

Table of Contents

Natural Sciences

Humanities

Social Sciences

Meet the Author:

Megan Shaeffer is an undergraduate student at UNC-Chapel Hill. She is a fourth-year student in the Biology and Hispanic Linguistics programs and has plans to graduate in May 2024. Megan's research experience includes biocurating in the Berg Lab, and has also completed various interdisciplinary research projects, combining knowledge in the sciences with those surrounding language history.

Genetic Analysis of the Immigration and Emigration of Basque Country Throughout History

Megan Shaeffer

Key terms: gene pool, haplotype, Pompe disease, immigration, emigration, novel variant

Abstract

It is widely agreed upon that Basque Country, an autochthonous region located in Northern Spain along the Bay of Biscaya, has withstood periods of isolation, supported by the linguistic isolation of the region's language, *Euskera*. Additional support for this claim comes from the history of invasion by various parties and the effects experienced by them. In this investigation, I seek to corroborate this history through an interdisciplinary analysis. I follow the migration patterns of the inhabitants of Basque Country by utilizing maps of historic migration patterns and comparing them with the documented genetic information of the people in the area. I do this by analyzing haplogroups DF27, Z195, Z220, Z278, and M153, as well as the Pompe disease variants c. 1579_1580del and c.1075+2T>C, the latter of which is a recently-discovered novel variant in the area. Through my analysis, I found that there is an overlap with haplogroup distribution and invasion patterns, as well as Pompe disease variant presence with recent global migration patterns.

Genetic Analysis of the Immigration and Emigration of Basque Country Throughout History

As a species, humans have evolved over millions of years. About 2.3 million years ago, the hominids called *Homo habilis* had evolved a larger skull to house brains capable of complex thought, as well as to develop simple tools. Approximately one million years ago, hominid species, such as the Homo erectus, began migrating out of Africa into what is now called Eurasia (Khan Academy). Different populations have developed throughout this migration, along with many years of geographic changes due to erosion and changes to the planet's climate,

These populations were unique in in their development of certain genetic features that contributed to their survival in their respective areas. For example, there are differences in the chemical compositions of those who live near the equator versus those who live away from the equator. This presents as eumelanin in individuals in the equatorial region, which helps block the harmful effects of UV radiation in greater quantities. Pheomelanin is responsible for augmenting the amount of UV radiation that is absorbed through the skin, which is responsible for the production of vitamin D in moderation (UCSB).

This variation in the favorable characteristics constitutes the gene pool, a set of all genes in any given population. These gene pools developed to be specific to a region and were maintained by the lack of contact between populations during prehistoric times and the development experienced over tens of thousands of years. This is what led humans to appear unique from one another. However, with advances in transportation and navigation, these different populations started to expand to neighboring populations and, with more major advancements, to other parts of the planet.

This has led to genetic traits that are characteristic of a specific region being seen in other populations. For example, the conquest of the Iberian Peninsula by the Vandals allowed certain characteristics seen in the gene pools of specific areas to appear in surrounding areas. This can be attributed to the movement of outside armies through these regions, and possibly the recruitment or conscription of new soldiers from these respective regions. These individuals would then reproduce with individuals in recently conquered areas, incorporating traits from their region's gene pool into the new one. Additionally, with the increase in globalization recently, there are also occurrences where certain genetic characteristics common in only one population suddenly appeared on the other side of the world. This would allow the incorporation of genes from one's respective gene pool to be introduced to to the region to which they have migrated.

These tendencies can be seen in individuals of Basque Country. Located along the northern border of Spain, along the Bay of Biscaya, this region of the country is home to an ethnically autochthonous community. Basque country has experienced periods of isolation throughout its history, which is reflected in its language, religion, and, as is being argued for here, certain heritable traits.

There are certain traits that, at one time, were unique to Basque Country, but they are now found in other areas of Europe, as demonstrated by Figures 1 and 2. These other European regions, such as the rest of the Iberian Peninsula, have also been conquered by external forces in the past.

The Basque gene pool has evolved to not only harbor unique traits and features to the population here but also to be inclusive of traits that have previously been seen outside of Europe. Additionally, consequential to globalization, these uniquely Basque traits

have recently been seen in other places in the world. In this essay, I will look at examples of genetic relationships between the Basque gene pool and other gene pools and relate them to patterns of immigration and emigration throughout history. Examples of genetic trends seen in these populations will be the appearance of specific haplogroups throughout the Iberian Peninsula, with some genetic variations of Pompe disease that have been seen in the Basque Country. This will be able to demonstrate the existence of a specific Basque gene pool, unique and separate from other gene pools in Europe and countries around the world.

A haplogroup is a specific population that shares a common ancestor through the paternal or maternal line (Rishishwar & Jordan 2020). A haplogroup, generally, is specific to a particular ancestry. For example, there is evidence to suggest the existence of a "Viking" haplogroup, R1a1 (Lall et al 2020). One's haplogroup is one of the ways scientists determine maternal or paternal ancestry. Paternal haplotypes are linked to the Y chromosome and are transmitted from the father to male offspring. Mitochondrial DNA, or mtDNA, is transmitted from mothers to male and female offspring. However, only female individuals will be able to pass on their mother's mtDNA. This is due to biologically male individuals having an X and Y chromosome, while biologically female individuals possess two X chromosomes. Because of this, female offspring cannot receive any DNA that would be present on a Y chromosome. Male offspring are able to receive mtDNA, but are not able to pass it on outside of rare occurrences, a process that is further discussed in Steven Z. DeLuca and Patrick H. O'Farrell's *Barriers to Male Transmission of Mitochondrial DNA in Sperm Development.*

This means that in lineages where males do not have male offspring, the Y chromosome haplotype of that individual's

ancestry will not be passed on to future generations. For example, if a man possessed the Scandinavian haplotype and had only female offspring, the offspring would not possess that haplotype because it is linked to the Y chromosome, which biological females would not possess. However, it is important to note that this does not mean that this individual will not be determined to be of Scandinavian ancestry in genome analysis because there are other ways to determine ancestry beyond the scope of this investigation.

A similar occurrence is seen with biological females. If a biological female were to give birth to a biological male, the male would receive the haplogroup of the mother in the form of mtDNA. For example, if the birthing parent possessed the Celtic haplogroup, all of their children would possess this in the form of mtDNA. However, as previously stated, only the biologically female offspring would be able to transmit this to future generations.

Haplogroup DF27 is a Y-linked haplogroup found almost exclusively in Europe. The highest concentration of this haplogroup is in the Iberian Peninsula, particularly in Basque Country (Solé-Morata et al., 2017). There is also a high concentration of haplogroup DF27 moving southwestward across the Iberian Peninsula toward the Strait of Gibraltar, which is represented in Figure 1 where red indicates the highest cluster of DF27, with a color gradient moving southwest through the Iberian Peninsula. Interestingly, this follows the path of the Vandal invasion of the Roman Empire in 409 A.D., leading them through the Basque Country and Spain to North Africa, as represented by the pink line in Figure 2 (Rush 2018). Because of this, it is reasonable to conclude that there is a causal relationship between the Vandal invasion route through the Roman Empire and the

concentration of individuals with haplogroup

DF27.

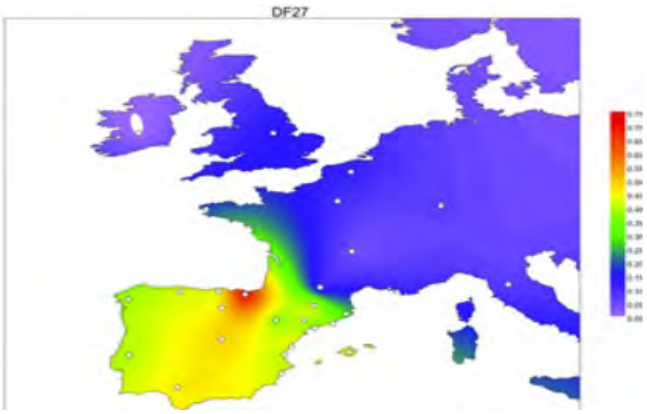


Fig. 1. Solé-Morata et al, *Contour map of DF27 in Iberian Peninsula*, 2017

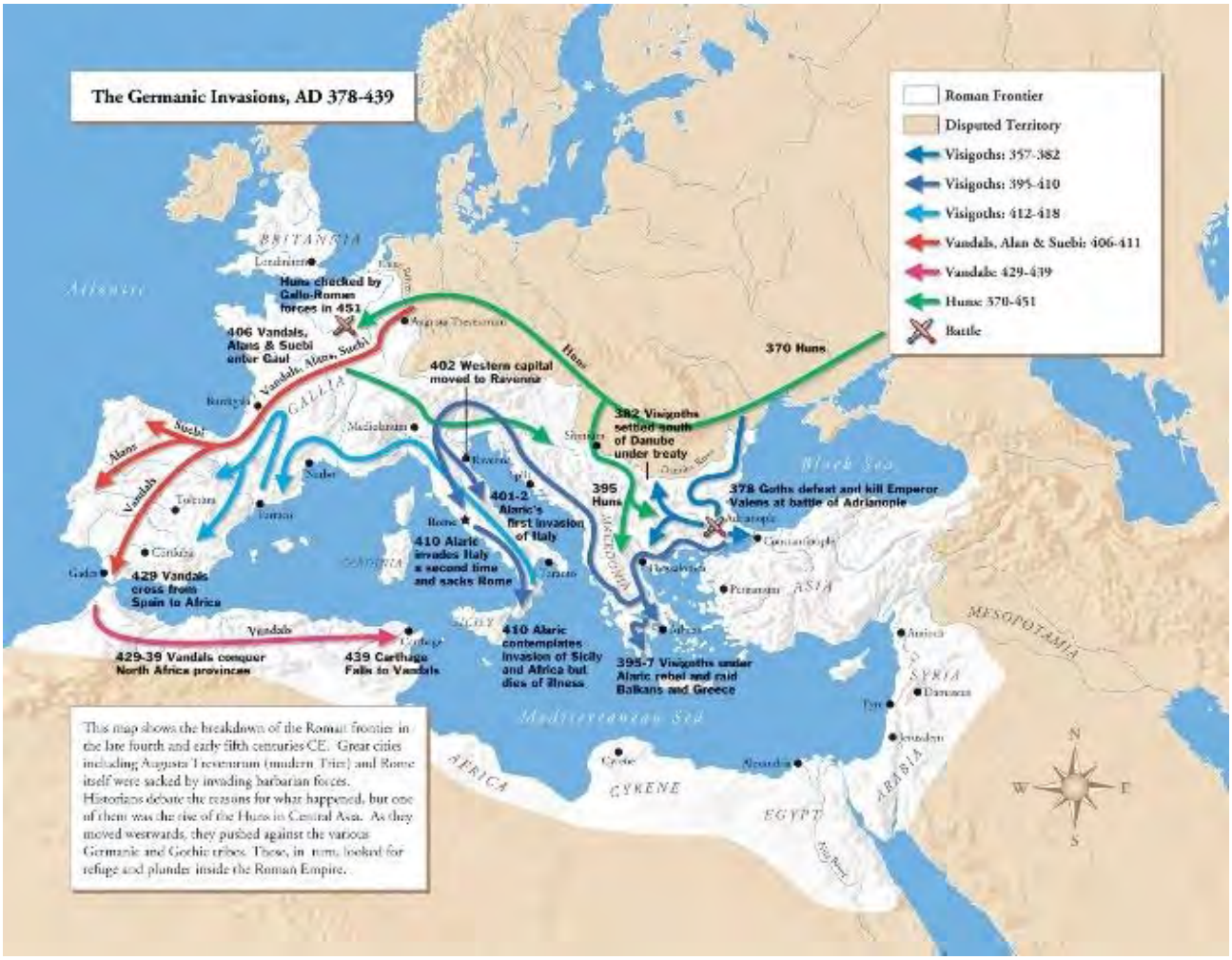


Fig. 2. Rush, *Tribal Migration and Invasion*, 2018

Based on the comparison between these two images, it is probable that DF27 originated in Basque Country. The high concentration of individuals with this haplogroup in this area helps to affirm the existence of a specific Basque gene pool. However, in addition to DF27, there are also other haplogroups from the Basque Country that have been seen in the rest of Europe. The haplogroups I will refer to are Z195, Z220, Z278 and M153. These haplogroups also have high concentrations in the Basque Country but, unlike the first haplogroup, they have interesting distributions in the Iberian Peninsula. While DF27 follows some kind of even gradient through the Iberian Peninsula, these haplogroups were concentrated in specific areas.

Solé-Morata et al. found that these haplotypes are related to DF27. They are the result of mutations that have occurred within populations with the DF27 haplotype, which in turn have then been transmitted to offspring from generation to generation. Figure 3 depicts the phylogenetic lineage tree of these haplogroups, representing the evolutionary descent of the new haplogroups from DF27, starting in the III-1 position in the tree. It can be assumed that these new haplogroups have evolved within the new populations and have been perpetuated in their gene pools. This explains the concentration of these new haplotypes in specific populations, as shown in Figure 4. This, again, is evidence of an existing Basque gene pool, as well as evidence of the movement of Basque descendants to neighboring populations.

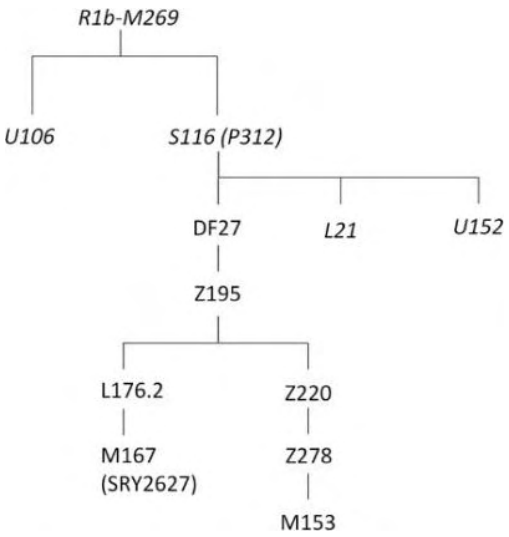


Fig. 3. Solé-Morata et. Al, *Simplified phylogenetic tree of haplogroup R1b-M269*, 2017

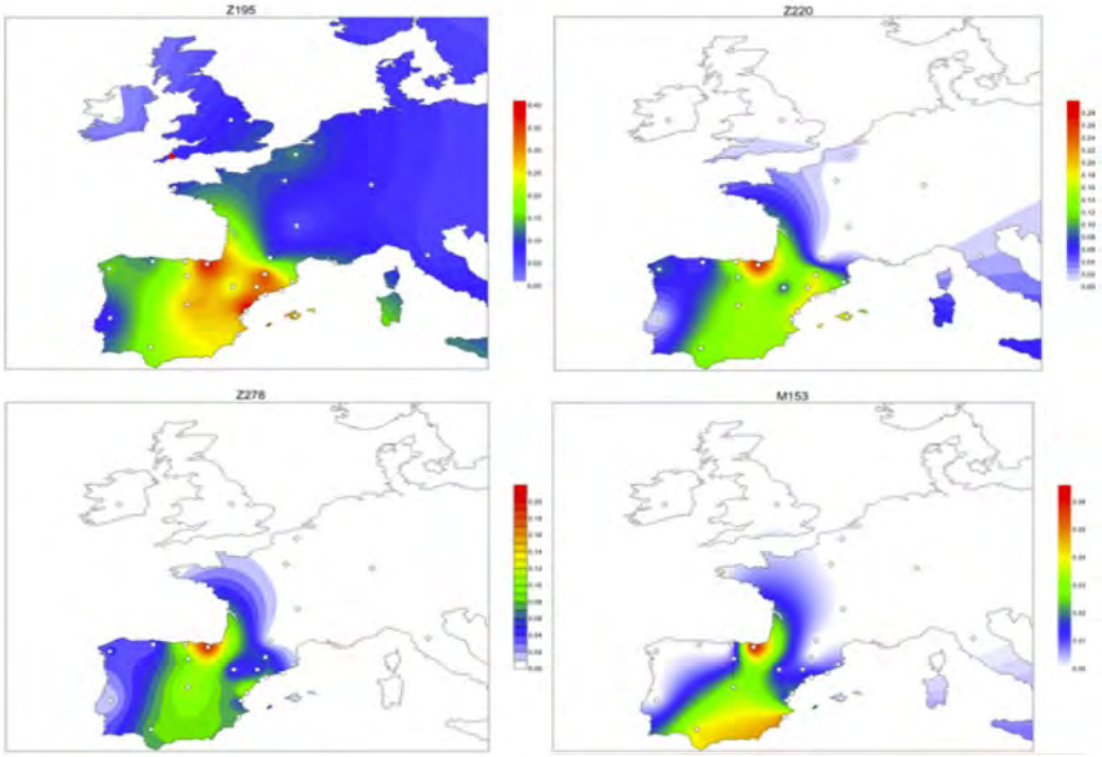


Fig. 4. Solé-Morata et. al, *Contour maps of Z195, Z220, Z278 and M153*, 2017

There are other ways to see the impact of genetics in a region; for example, through genetic conditions and diseases. Genetic diseases are inherited by an individual who, in the majority of cases, was born from parents who both passed on the variant allele to the offspring. Oftentimes, there are cases of the genetic disorder occurring in each of

the parents' families. For the purposes of this paper, I will refer to autosomal disorders. The majority of known autosomal disorders are recessive, meaning the individual must receive the genetic information containing the disease from both parents. For example, an individual who inherited genetic information that encodes for a certain genetic disorder from their biological mother but not from their biological father will not express the genetic disorder in question. However, if the offspring were to receive the disordered genes from both parents, they would express the genetic disorder.

Pompe disease is a rare genetic disorder that causes progressive weakness in the heart and skeletal muscles (Dasouki et al., 2014). It is an autosomal recessive disorder, meaning that both of the parents of an affected individual would have to transmit a copy of the genetic information encoding for Pompe disease. There are several types of Pompe disease: classic infantile-onset, non-classic infantile-onset, and late-onset. As the names suggest, these different types of Pompe disease arise at different periods of an individual's life. In addition, they have different severities, with infantile-onset being the most severe and late-onset generally being the least severe (Leslia & Bailey 2017).

As with haplogroups, there are certain mutational variants of Pompe disease that are unique to specific regions of the world. For example, the c.1579_1580del variant is a variant commonly seen in Asia. Recorded cases from 2018 and 2019 reported two separate individuals, both in Korea, with this variant (Ko et al., 2018; Kim et al., 2019). These individuals are the only ones in the world that had been documented with this variant at the time. For this reason, we can consider this variant to be characteristic of Asia. This is significant because, assuming that this gene originates from an Asian gene pool, this means that there would have to have been some form of migration of this

variant from Asia into a separate gene pool, should it be documented in that location outside of Asia.

Interestingly, in the Basque Country, variant c.1579_1580del has emerged. This individual, who is of Nepalese ancestry, experiences infantile-onset Pompe disease and is reported to be heterozygous with the variant, meaning that the individual received the c.1579_1580del allele from one parent and another allele that causes Pompe disease from the other (De las Heras, 2021). In this situation, the variant in question was received from the mother, while another recessive variant was received from the father. This supports the idea of genetic trends that reflect the journey of individuals between gene pools, considering that there is an extremely low probability of random mutations occurring in the genome. The estimated number of genetic units that are changed is approximately 1 in 10,000,000,000,000 (Carlin 2011).

Evidence of this migration from Asia to the Basque Country has been recorded the past two and a half decades. In 2013, the Chinese population in Basque Country reached 11,635 people (Ikusipegi), and is the largest Asian population in Basque Country. This number has been steadily increasing since 1998 and is likely to continue to rise. It is possible that some of these individuals are carriers of the c.1579_1580del Pompe disease variant, due to proximity with Korea, and had simply not been recorded until recent times. It is also possible that this variant is not specific to only Korea, but is characteristic to a larger Asian population. Further research will need to be done on this subject in order to gain a more complete understanding of the nature of c.1579_1580del.

In addition to the c.1579_1580del variant, there is a new variant, referred to as a novel variant, that has been found in the

Basque Country and has never been recorded in any scientific literature database. This variant, c.1075+2T>C, is present in another individual from the Basque Country. Upon confirmation with Dr. Javier De las Heras by email, the family of this individual is originally from the Basque Country, which means that this variant has most likely arisen in a Basque gene pool.

It is also likely that other Basque families have this variant. We can assume this statement is correct because variants, in general, are already established in a family's lineage, due to the very low chance of a mutation occurring randomly. This can be proven due to emerging scientific practices that allow for families to easily test for genetic variants. However, because Pompe disease is a recessive disorder, an individual who were to possess c.1075+2T>C would be required to have a second variant that also causes Pompe disease in order to express the disorder. It is possible that there are many individuals in Basque Country who are carriers for this variant but have not undergone genetic testing. Therefore, it cannot be proven at this moment. However, this is an area for future research.

This is important for arguing the existence of a unique Basque gene pool, taking into account that this gene has not been found in other countries. Therefore, it is unlikely that individuals migrating from other gene pools may have brought this gene to Basque Country. These cases of Pompe disease support the idea that there are specific Basque Country genes, specifically the c.1075+2T>C variant, which form a gene pool. Therefore, there may be genes specific to other locations, such as the c.1579_1580del variant, that can be introduced into the Basque Country gene pool and begin to increase in prevalence.

Based on these factors, there is a unique Basque gene pool that exists in Europe. There is evidence of Basque descent

in other parts of the Iberian Peninsula based on the Vandal conquest trail. This is significant because, for this to happen, there had to be strong concentrations of a specific gene that could be seen in a multitude of individuals in the Basque Country. Said gene would then be able to be spread to neighboring regions through an increase in contact due to being under the same conquering group.

Something similar can be said about other gene pools. If an existing separate gene pool has a strong prevalence of a specific gene, then the probability of a seemingly random person possessing that gene is very high. This means that if an individual were to bring a gene unique to their own region to a new area, such as from Asia to the Basque Country, there would now be appearances of that gene in the new region.

Because the existence of a Basque gene pool has now been established, based on haplogroup DF27 and the c.1075+2T>C variant of Pompe disease, we can now see how it relates to human migration patterns to and from the Basque Country throughout history. As recently mentioned, the movement of haplogroup DF27 across the Iberian Peninsula reflects the conquest of the Vandals in that area after contact with the Basque Country. The appearance of new haplogroups throughout that area, such as Z195, Z220, Z278, and M153, are simply the result of several mutations, each occurring in unique regions, which have since been perpetuated and passed onto future generations in these areas. In addition, the appearance of a variant of Pompe disease in the Basque Country that was once reported mainly in Asia reflects the increase in Asian immigration to the Basque Country in the last two and a half decades.

Changes in genetic trends can be used as a strong indicator of human migration patterns. As seen with the history of migration in the Basque Country in

combination with the spread of certain genes in the Iberian Peninsula, we can now better understand the relationship between gene flow between populations and possible external influence.

Works Cited

Carlin, Joel L. Nature News, Nature Publishing Group, 2011, https://www.nature.com/scitable/knowledge/library/mutations-are-the-raw-materials-of-evolution-17395346/#:~:text=In%20point%20mutations%2C%20one%20base,10%2C000%2C000%2C000)%20base%20pair%20is%20changed.

Dasouki, M., Jawdat, O., Almadhoun, O., Pasnoor, M., McVey, A. L., Abuzinadah, A., Herbelin, L., Barohn, R. J., & Dimachkie, M. M. (2014, August). Pompe Disease: Literature Review and Case Series. Neurological Clinics.

De las Heras, Javier. Email to Megan Shaeffer. 2 February 2023.

De las Heras, Javier, et al. “Importance of Timely Treatment Initiation in Infantile-Onset Pompe Disease, a Single-Centre Experience”. Niños (Basilea, Suiza), Biblioteca Nacional de Medicina de EE. UU., 9 de noviembre de 2021, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8620435/.

DeLuca, S. Z., & O’Farrell, P. H. (2012, March 12). *Barriers to male transmission of mitochondrial DNA in sperm development.* Science Direct. https://www.sciencedirect.com/science/article/pii/S1534580712000020

“First Humans: Homo sapiens and early human migration.” Khan Academy, Khan Academy, https://www.khanacademy.org/humanities/world-history/world-history-beginnings/origin-humans-early-soccieties/a/where-did-humans-come-from.

Kim, Min-Sun, et al. “Clinical and molecular characterization of Korean children with infantile and late-onset Pompe disease: 10 years of experience with enzyme replacement therapy at a single center”. Korean Journal of Pediatrics, Biblioteca Nacional de Medicina de EE. UU., junio de 2019, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6584236/.

Ko, Jung Min, et al. " A New Integrated Newborn Screening Workflow Can Provide a Shortcut to Differential Diagnosis and Confirmation of Inherited Metabolic Diseases". Yonsei Medical Journal, Biblioteca Nacional de Medicina de EE. UU., 1 de julio de 2018, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5990675/.

Lall, G. M., Larmuseau, M. H. D., Wetton, J. H., Batini, C., Hallast, P., Huszar, T. I., Zadik, D., Aase, S., Baker, T., Balaresque, P., Bodmer, W., Børglum, A. D., de Knijff, P., Dunn, H., Harding, S. E., Løvvik, H., Dupuy, B. M., Pamjav, H., Tillmar, A. O., … Jobling, M. A. (2020, November 2). Subdividing Y-chromosome haplogroup r1a1 reveals Norse Viking dispersal lineages in Britain. Nature News. https://www.nature.com/articles/s41431-020-00747-z

Leslie, N., & Bailey, L. (2017, May 11). Pompe Disease. Europe PMC. https://europepmc.org/article/NBK/nbk1261

“Población Asiática En La CAPV.” Ikusipegi, https://www.ikuspegi.eus/documentos/panoramicas/es/panoramica52cas.pdf.

"Pompe disease". MedlinePlus, National Library of Medicine of the United States, https://medlineplus.gov/genetics/condition/pompe-disease/.

Rishishwar, Lavanya y I King Jordan. “What are Haplogroups? Living DNA explain.” Living DNA, 19 de marzo de 2020, https://livingdna.com/blog/haplogroups-explained.

Rush, I. (2018). Retrieved from https://courses.wccnet.edu/~jrush/121outline10.html

Solé-Morata, N., García-Fernández, C., Urasin, V., Bekada, A., Fadhlaoui-Zid, K., Zalloua, P., . . . Calafell, F. (2017, November 21). Whole Y-chromosome sequences reveal an extremely recent origin of the most common North African paternal lineage E-M183 (M81). Retrieved April 10, 2023, from https://www.nature.com/articles/s41598-017-16271-y

“Why do people have different skin colors?” Línea científica de UCSB, http://scienceline.ucsb.edu/getkey.php?key=3208.

Meet the Author:

I am Venkata Mantri, an undergraduate Biology student at UNC. I got into research at first as a method of self-learning about Biological Aging. Through some basic research I found a deep interest in biological aging, in both how it is caused by our regular metabolism and in what can be done to stop or reverse aging, because, at least to me, it seems to be a major problem that we now have a chance of tackling. I realized that I would have to learn a lot about a certain topic if I wanted to write a paper on it, and so this paper acts as both a personal project and something like the culmination of my biological aging knowledge, at least up until now. I am currently working in the Griffith lab where we work on telomeres!

A Review of the Interactions Between the Causes of Aging

Venkata Sivasaisumarishi Mantri

Independent Research

Abstract

As the elderly population increases, there has been a rise in the number of age-related diseases. Instead of combating these diseases individually, it would be more effective to target the components of aging itself. Recent advancements in aging research have allowed for the discovery of multiple possible causes of aging which, if treated, could fight aging. These can be summarized as genetic/epigenetic damage, loss of proteostasis, mitochondrial dysfunction, senescent cells, altered intercellular communication, and stem cell exhaustion. These causes are very interconnected and can each interact with the others in a multitude of ways. These interactions build a complex network between the different causes of aging and show how they promote each other. This review lays out a current understanding of these interactions and their effects.

Acknowledgements:

Keywords: Aging; Mitochondrial Dysfunction; Proteostasis; Telomere; Genome Instability;

I. Introduction

Recent advances in medicine, public health, and education have allowed for an increased average human lifespan. As a result of this, the elderly population has grown substantially and the detrimental effects of aging have become much more prevalent. This has led to the increased rates of age-related diseases, such as cardiovascular disease, cancer, osteoporosis, neurodegenerative diseases, and diabetes mellitus [1]. While the diseases caused by aging can be dealt with independently, recent advancements in understanding and combating the aging process have made it possible to fight aging itself rather than treating the diseases one by one.

Researchers have found biological processes that are associated with aging and could be interpreted as the significant drivers of aging. These processes have been laid out relatively recently in many scientific sources, primarily “Ending Aging” and “The Hallmarks of Aging” [2,3]. The causes of aging are generally considered to be genetic/epigenetic damage, loss of proteostasis, mitochondrial dysfunction, senescent cells, altered intercellular communication, and stem cell exhaustion. While there has been substantial progress in aging research and multiple treatments/therapies being researched to combat the causes of aging in various ways, aging is a multifaceted and synergistic combination of these processes that requires a comprehensive understanding to combat effectively. The interactions between these causes are numerous and complex. This paper outlines the current understanding of their interactions.

II. The Primary Causes: Genetic / Epigenetic Damage, Loss of Proteostasis, and Mitochondrial Dysfunction

A. What are these causes?

Among the various factors that cause aging, genetic/epigenetic damage, loss of proteostasis, and mitochondrial dysfunction can be seen as the root causes of aging. These processes influence both downstream causes as well as each other in significant ways.

Genetic and Epigenetic Damage

DNA is the genetic information within every cell that contains the instructions for how to make proteins, and as such, it is vitally important for cell survival. It has been estimated that approximately 10^4–10^5 DNA lesions are generated in each mammalian genome per day. However, since DNA is an extremely important macromolecule, each cell makes enormous investments in DNA repair to create an elaborate genomic maintenance system to respond to such damage [4]. There are multiple different types of DNA repair, each responsible for different types of damage, and if these processes themselves, or the genes responsible for these repair processes, become deficient they can lead to cancers, apoptosis, and other problems [5, 6, 7, 8]. A key defense against these cancers are telomeres.

Telomeres are caps at the end of chromosomes which consist of non-coding double-stranded repeats of guanine-rich tandem DNA sequences which can fold back into the double-stranded telomere helix, forming a large ‘T-loop’ [9]. As cells divide, their telomeres gradually shorten because most cells do not have the machinery to fully replicate their telomeres, and as a result the daughter cells receive chromosomes that have shorter telomeres with each division [3]. As telomeres gradually shorten the cell loses the ability to divide, and when the critical length is reached DNA repair and cell cycle checkpoint mechanisms are triggered improperly, resulting in chromosomal fusions, cell cycle arrest, senescence, and/or apoptosis [10]. However,

some cells can overcome the replication limit posed by telomeres such as germline cells, stem cells, and cancer cells which allows them to divide continuously. This ability is usually achieved through the transcription and translation of telomerase, which allows these cells to regenerate their telomeres [11].

Outside the physical structure of DNA, epigenetics represents the reversible heritable mechanisms that change phenotype or gene expression without any alteration to the underlying DNA sequence. The genome carries genetic information and the epigenome is responsible for the functional use and stability of that information by connecting the genotype to the phenotype. The epigenome could also serve as an explanation for why aging is experienced differently between two genetically identical individuals since the epigenome can change based on external or internal influences [12]. Epigenetic changes such as DNA methylation, histone modifications, chromatin remodeling, and non-coding RNAs have all been found to be related to age-associated phenotypes [13]. Additionally, epigenetic remodeling has been used to “reverse” the age of somatic cells, turning them into induced pluripotent stem cells.

Loss of Proteostasis

Proteostasis, or protein homeostasis, is composed of multiple pathways which are responsible for protein synthesis, folding, trafficking, aggregation, disaggregation, and degradation [14]. Proteostasis is maintained by the Proteostasis Network (PN), a complex system of molecular chaperones, proteolytic pathways, and protein degradation machines that work both independently and interconnectedly to make sure proteins work properly and deal with protein misfolding [15, 16]. In order to maintain proteostasis, the three parts of the PN, protein synthesis and folding,

maintenance, and degradation, must function properly. Major aspects of the PN, especially during degradation, include molecular chaperones, the ubiquitin-proteasome system, and the lysosome autophagy systems. Molecular chaperones are proteins, often heat-shock proteins such as HSC70, that assist in the folding, assembly, conformational maintenance, or regulation of another protein without becoming part of its final structure [17, 18]. Chaperones can assist proteins to either refold or degrade depending on the situation and can be affected by factors such as chaperone availability and cellular ATP content, as some chaperones are ATP dependent [19]. When protein functionality cannot be restored, chaperones may help direct the proteins toward degradation pathways, which are the ubiquitin proteasome system (UPS) and the lysosome autophagy system (LAS). In contrast to the UPS, the LAS is restricted to the cytoplasm but degrades a much wider spectrum of substrates. While both the UPS and LAS are capable of degrading soluble misfolded proteins, the UPS mostly degrades single, unfolded polypeptides which can fit the narrow channel of the proteasome whereas the LAS deals with larger, cytosolic structures such as protein complexes, cellular aggregates, organelles, and pathogens [20, 21].

Mitochondrial Dysfunction

Mitochondria are double membrane enclosed organelles that are involved in multiple cellular processes such as the citric acid cycle, oxidative phosphorylation (OXPHOS), ATP production, apoptosis, β-oxidation of fatty acids, and iron-sulfur cluster synthesis [22]. Unlike many other organelles, mitochondria contain their own mitochondrial DNA (mtDNA) which, in humans, encodes 13 proteins, 22 tRNAs, and 2 rRNAs. Although the mtDNA only encodes around 1% of the mitochondrial

proteome, these 13 proteins are vital for proper mitochondrial functioning, as they are all critical components of the OXPHOS complexes [22]. Their main function is the creation of ATP through OXPHOS, which is vital to multiple processes within a cell. As aging progresses mitochondrial function becomes impaired, and this usually involves altered mitochondrial respiration, decreased energy production in the form of adenosine triphosphate (ATP), and widespread changes in metabolites associated with mitochondrial function [23, 24, 25]. This decline in mitochondrial functioning with age can be dubbed mitochondrial dysfunction.

B. Their interactions.

1. Genetic/epigenetic damage and loss of proteostasis

Genetic/epigenetic damage and loss of proteostasis have a very direct connection in that DNA damage or incorrect repair of this damage can occur in genes and result in misfolded proteins, leading to the loss of proteostasis, since a major cause of the loss of proteostasis is the build up of misfolded proteins. Observations in cellular and organismal models show that chronic production of misfolded and aggregated proteins compromises the proper functioning of the PN, including its capacity to fold proteins, clear misfolded proteins, and respond to stress by upregulating the PN [26]. Usually, the accumulation of misfolded proteins during times of stress, such as heat stress, creates a rapid PN response to restore proteostasis, but this doesn't seem to happen during chronic production of misfolded proteins [26]. These misfolded proteins can aggregate into huge masses which become much harder to degrade and can ultimately result in lipofuscin. Lipofuscin is a fluorescent complex mixture of multiple cross-linked oxidized molecules that aggregate and are not degradable by

lysosomes [27]. This causes them to aggregate in the lysosomes, which leads to significantly reduced lysosomal functioning and decreased lysosomal enzyme activity [28].

While damage to DNA results in declining proteostasis, this also seems to work in reverse as this decline in proteostasis can also result in further DNA damage and even contribute to cancer. Damaging the DNA that affects proteins involved in DNA repair results in decline of both proteostasis and the cell's ability to maintain genome integrity. This is especially the case in cancers as DNA damage is a hallmark of cancer, with some proposing that all cancers are connected to defects in DNA repair and around 90% of all cancers have a mutation of reactivate telomerase [29, 30].

How the epigenome and proteostasis interact is somewhat similar, as changes to the epigenome impacts the expression of certain genes, which affects the production of certain proteins and disrupts proteostasis. For example, hypermethylation of LC3 and autophagy-related gene 5 (Atg5) promoter regions has been observed in the macrophages of old mice. This epigenetic modification downregulates protein levels and promotes the decline of autophagy [31, 32]. Additionally, a balance between protein acetylation and methylation can be presumed to be important for autophagy, because these processes can impact gene expression [31]. Beyond these interactions, it also seems that both the unfolded protein response (UPR) and DNA damage response (DDR) interact with each other in significant ways, though not too much is known in this regard. For instance, in S. cerevisiae, exogenous expression of mammalian X-box binding protein 1, which is a major player in the endoplasmic reticulum proteostasis, was found to play a role in non-homologous end joining and double strand break repair

pathways through the regulation of H4 acetylation [33]. Moreover, in a panel of cancer cell lines, genotoxic drugs were found to promote changes in ER structure through the transcriptional activation of p53, a tumor suppressor and major part of the DDR. This resulted in upregulated expression of receptor expression-enhancing proteins 1 and 2 and upregulated p53-induced gene 8, which are three important endoplasmic reticulum shaping proteins [33, 34].

2. *Genetic/epigenetic damage and mitochondrial dysfunction*

Interactions between genetic/epigenetic damage and mitochondrial dysfunctions are numerous, but one major way they interact is through oxidative stress caused by reactive oxygen species (ROS). There are many sources of intracellular ROS in mammals, but the mitochondrial electron transport system has been recognized to be responsible for almost 90 % of the total ROS produced in the cell with the two major sites of ROS generation believed to be complex I and complex III in their respiratory chain [35-38]. ROS can damage the genome and can directly produce oxidatively damaged bases, abasic (AP) sites, and single-strand breaks. This can cause multiple problems, such as alteration of the double helix structure or mutations [4]. Additionally, ROS can damage other cellular macromolecules and produce intracellular genotoxins. For example, lipid peroxidation can create malondialdehyde, which can react with DNA to generate bulky adducts or promote the formation of interstrand crosslinks. This can create structural alterations in DNA and inhibit the replication or transcription machinery [4, 39].

Beyond the damage to the DNA, this creates other problems. At the beginning of DNA repair, poly ADP-ribose polymerases (PARPs) are activated and play key roles in the cellular response to DNA damage [40]. These PARPs require ATP and Nicotinamide adenine dinucleotide (NAD+) to function properly [40]. In fact, it has been found that concentrations of NAD+ and ATP decrease rapidly after activation of PARP1, and chronic activation of PARP1 can result in sustained depletion of NAD+ and ATP [40, 41, 42]. This loss of NAD+ is important as NAD+ is heavily used by mitochondria in OXPHOS to produce ATP and is used for the functioning of sirtuins. In support of this, it has been shown that mice with mutations in PARP1 or treated with PARP inhibitors had improved mitochondrial function and organismal fitness, and worms with PARP inhibition seem to have extended lifespan [25, 43-45]. However, it is important to consider that PARP is partly responsible for DNA repair and long term inhibition of PARP may lead to increased amounts of DNA damage along with increased susceptibility to cancers. This could also be a reason why NAD+ supplementation has been found to be beneficial to health and lifespan. NAD+ supplementation has extended lifespan and improved healthspan in yeast, flies, worms, and mice, as shown by improved mitochondrial health, muscle strength, and motor function [46-49]. However, this is probably connected to the decline in NAD+ levels during aging as it has been reported that levels of NAD+ decrease in mice, C elegans, and humans during aging. By the time a mouse or human is middle-aged, levels of NAD+ have fallen to half of youthful levels, which can result in loss of PARP, mitochondria, and sirtuin activity [47, 50].

Sirtuins are another way DNA and mitochondria interact. Sirtuins are NAD+

dependent proteins that are important in genetic, epigenetic, and mitochondrial health as they are involved in the DDR, chromatin modeling, transcription, metabolism, and mitochondrial functioning [25, 51]. Mammals have 7 sirtuins (SIRT) and sirtuins 3-5 (SIRT3-5) localize in the mitochondria and are dubbed mitochondrial sirtuins, but SIRT1 has also been shown to affect mitochondria. SIRT3 seems to be the largest player in mitochondria in comparison to SIRT4 and SIRT5 and has been found to be involved in ROS balance, OXPHOS, fatty acid metabolism, ketone body metabolism, amino acid catabolism, the citric acid cycle, the mitochondrial UPR, mtDNA functions, and mitochondrial translation while SIRT4 and SIRT5 seem to interact with a narrower set of mitochondrial functions [52]. Along with these functions, SIRT3 also localizes at DNA break sites during times of nuclear DNA damage, where it is required for 53BP1 foci formation by deacetylating H3K56 residue and allowing for non-homologous end joining [53, 54]. This could also be a reason that PARP inhibition improves mitochondrial function. PARP is involved in repairing DNA break sites, and while it is unclear whether PARP can regulate SIRT3 activity, cisplatin-damaged renal cells exhibited reduced SIRT3 levels that were improved with treatment by a PARP inhibitor, which hints at possible interaction between PARP and SIRT3 [55, 56]. Along with influencing DNA and mitochondria, sirtuins also play a role in maintaining the epigenome. For example, SIRT3 deacetylates all the electron transport chain complexes to promote efficient electron transport and maximize ATP production in the mitochondria, but in a similar fashion SIRT3 is also involved in H4K16ac deacetylation when transported to the nucleus under specific conditions [52, 57, 58].

Beyond sirtuins, a major way that mitochondria affect the epigenome is through substrate availability. Histone acetyltransferases (HATs) and histone deacetylases (HDACs) are enzymes responsible for adding and removing acetyl groups on histone tails and are dependent on acetyl-CoA and ATP for proper functioning. However, acetyl-CoA and ATP are both heavily reliant on mitochondrial processes such as OXPHOS and fatty acid oxidation [59]. Mitochondrial functioning also plays a role in methylation of both DNA and histones. Histone methyltransferase (HMT) enzymes are responsible for creating methyl marks while histone demethylases (HDMs) remove methyl marks, and unlike acetylation histone methylation can both activate or repress transcription by adding up to 2 or 3 methyl groups to a lysine or arginine residue on a histone [59]. Mitochondria have involvement in histone methylation because HMTs use S-adenosyl methionine (SAM) as a precursor to methylate histones, and even though SAM is synthesized through the methionine–homocysteine cycle in the cytosol, this cycle depends on the folate cycle and ATP, which are dependant on mitochondria [59-61]. Some HMTs have also been found to modulate mitochondrial function. For example, inhibition of lysine methyltransferase SETD7 (SET7/9) has been found to promote mitochondrial biogenesis and activate an antioxidant response [59]. For epigenetic modification in the form of DNA methylation, individual nucleotides in the DNA are altered, with the most common alteration being the 5-methylcytosine (5mC), although other DNA modifications such as N6-methyladenine (6mA) have also been identified [62, 63]. In mammals, three DNA methyltransferases (DNMTs) that create 5mC methylation have been identified: DNMT1, DNMT3A, and DNMT3B. Similarly to HMTs, DNMTs use

SAM as a methyl donor to create 5mCs, and so mitochondria can affect DNA methylation through a similar pathway as histone methylation. However, in contrast to HMTs, DNMTs directly affect mtDNA through DNA methylation of the promoter or intronic regions. For instance, two enzymes involved with lipid oxidation (CPT1a and ACACA) directly affect the mtDNA epigenome [59].

Finally, the epigenome is also affected by ROS. Aside from the possibility that ROS-induced DNA damage happens or is repaired improperly at genes that code for proteins involved in epigenetic maintenance, almost any damage to DNA can negatively impact the epigenome. After DNA damage occurs, chromatin must be remodeled and moved to allow the DNA repair machinery to access the DNA damage, which can result in epigenetic changes. Beyond the movement of nucleosomes, histones are post-translationally modified at sites of DNA damage to aid in DNA damage signaling along with the recruitment and retention of DNA repair factors [64, 65]. After DNA damage is repaired, the epigenetic alterations are usually restored by repositioning nucleosomes and returning the epigenetic code back to its original state. However, if the epigenetic alterations are not removed correctly during the DNA damage repair process, or the cell is in a setting of chronic inflammation or chronic toxicant exposure, or there is repetitive DNA damage, these epigenetic changes can become permanent and change the expression of genes [64]. When talking about telomeres, ROS especially damages the telomeres and can cause telomere shortening. In fact, it has been found that oxidative damage is repaired less effectively in telomeres than elsewhere in the chromosome [66]. This makes telomeres especially vulnerable in regard to mitochondrial dysfunction.

3. *Mitochondrial dysfunction and loss of proteostasis*

Before the interactions between mitochondrial dysfunction and loss of proteostasis are laid out, it should first be noted that many typical cytosolic proteostasis mechanisms, such as proteasomes and heat shock proteins, do not function in the mitochondria.The mitochondria has its own chaperones, proteases, and other factors for quality control, such as mitochondrial Hsp70 (mtHsp70) and LonP. Many of these aid in protein folding and removal of damaged proteins since most of the proteins in the mitochondria are imported after being encoded by the nucleus. Moreover, mitochondria have a high protein turnover rate and are exposed to significant oxidative stress, which can alter proteins [67].

The ways that mitochondria and cellular proteostasis interact are numerous, but one is through the UPS. The UPS is greatly responsible for the ubiquitination and degradation of misfolded or unfolded proteins, and so it directly influences which proteins reach the mitochondria. This has been shown in cases of acute proteasomal inhibition, where it led to increases in subunits of respiratory complexes I, II, and IV, as well as the F1-F0-ATPase, and this indicates that many of these proteins are quickly degraded by the UPS before they reach the mitochondria [68, 69]. Additionally, it has been found that inhibition of proteostasis leads to a significant increase in levels of ubiquitinated proteins in the inner mitochondrial membrane, which suggests that ubiquitinated proteins are being imported into mitochondria when they should have been degraded. Proteostasis inhibition has also been associated with increased levels of

proteins being imported into the mitochondria and aggregation of respiratory complex subunits in the cytosol. All of these signs indicate a lack of degradation [5, 69-71]. The UPS is also involved in the ubiquitination and degradation of accessible proteins in the outer mitochondrial membrane (OMM) and proteins within the intermembrane space can actually be exported through the TOM complex, allowing for them to be disposed through proteasomal degradation [69, 72]. If the UPS operates well, proper proteins are transported to the mitochondria and are folded or disposed and transported properly by the mitochondrial proteome. However, when levels of misfolded proteins extend beyond the capacity of the resident chaperones and proteases, the mitochondrial unfolded protein response (UPRmt) is activated. An increase in misfolded proteins can happen due to multiple causes, including increased import of proteins from the cytosol. Misfolded proteins can aggregate in the cytosol and be transported to the mitochondria for degradation. This transport increases during times of stress when there are large amounts of protein that can aggregate. This allows the cell to regain control of protein aggregates in the cytosol while degrading proteins in the mitochondria, but transport can also happen during the absence of this stress, such as when certain cytosolic proteins with high structural instability are present [73, 74]. This could play a role in Alzheimer's Disease (AD) as the induction of the UPRmt is observed in various models of AD and amyloid β, a hallmark of AD, has been found to be transported into the mitochondrial matrix via the TOM complex [69, 75, 76]. While it's controversial that amyloid β is imported into mitochondria, amyloid β has been found to interfere with the mitochondrial import of proteins, and this may be because the amyloid β being

transported has already formed into aggregates and is disrupting the transport of proteins [69, 77]. While the import of proteins can usually be handled by the mitochondrial quality control machinery, excessive import can become toxic to mitochondria and can contribute to mitochondrial dysfunction.

Dysfunctional mitochondria cause a host of problems for the cell if not addressed, which is when autophagy takes place. Mitochondrial autophagy, or mitophagy, involves the degradation of dysfunctional mitochondria through the use of lysosomes and has been reported to be important for selective targeting and removal of dysfunctional mitochondria [22, 78]. Elimination of mitochondria through mitophagy is strongly connected with mitochondrial biogenesis in order to balance the removal and creation of mitochondria. This is supported by how Parkin, a significant player in causing mitophagy, simultaneously induces mitochondrial biogenesis through other mechanisms [69]. This is a major way that proteostasis and mitochondrial dysfunction interact since mitophagy and mitochondrial biogenesis try to ensure that the cell contains enough functioning mitochondria and eliminates dysfunctional ones. However, because dysfunctional mitochondria can damage proteostasis in multiple ways, mitophagy and proteostasis can be decreased, which then leads to more dysfunctional mitochondria not being removed. This creates a vicious cycle.

One way mitochondrial dysfunction can negatively impact proteostasis is by contributing to poor cell energetics. Poor cell energetics are common in old organisms and affect the overall availability of ATP in the cell, which can disrupt proper chaperone functioning because some chaperones are ATP dependent [19, 79]. This leads to reduced disposal of modified proteins,

which is worsened by dysfunctional mitochondria as they are also known to produce ROS which can oxidatively damage proteins. These protein modifications can further interfere with normal chaperone functioning [19]. Additionally, as stated previously, observations in cellular and organismal models show that chronic production of misfolded and aggregated proteins compromises proper functioning of the PN, and while accumulation of misfolded proteins during times of stress create a rapid PN response to restore proteostasis, this does not occur when there is a chronic production of misfolded proteins [26]. Some of these proteins may also undergo glycation and become advanced glycation endproducts, which can cause a variety of other problems, many of which are also age-related [80]. The production of ROS also contributes to the formation of undegradable intracellular substances such as lipofuscin, which heavily impacts proteostasis. This leads to reduced autophagy, which can then allow for further lipofuscin formation and aggregation of lipofuscin in the cytosol and lysosomes [27]. Additionally, mitochondrial dysfunction contributes to loss of proteostasis through containing ATP-synthase subunit-c, which appears to be the main component of lipofuscin in multiple neuronal ceroid lipofuscinosis diseases [27]. As all of these effects disrupt proteostasis and autophagy, mitophagy is decreased, leading to decreased mitochondrial biogenesis since, as stated above, the two are closely connected. As a result, dysfunctional mitochondria continue to survive while new healthy mitochondria are not created.

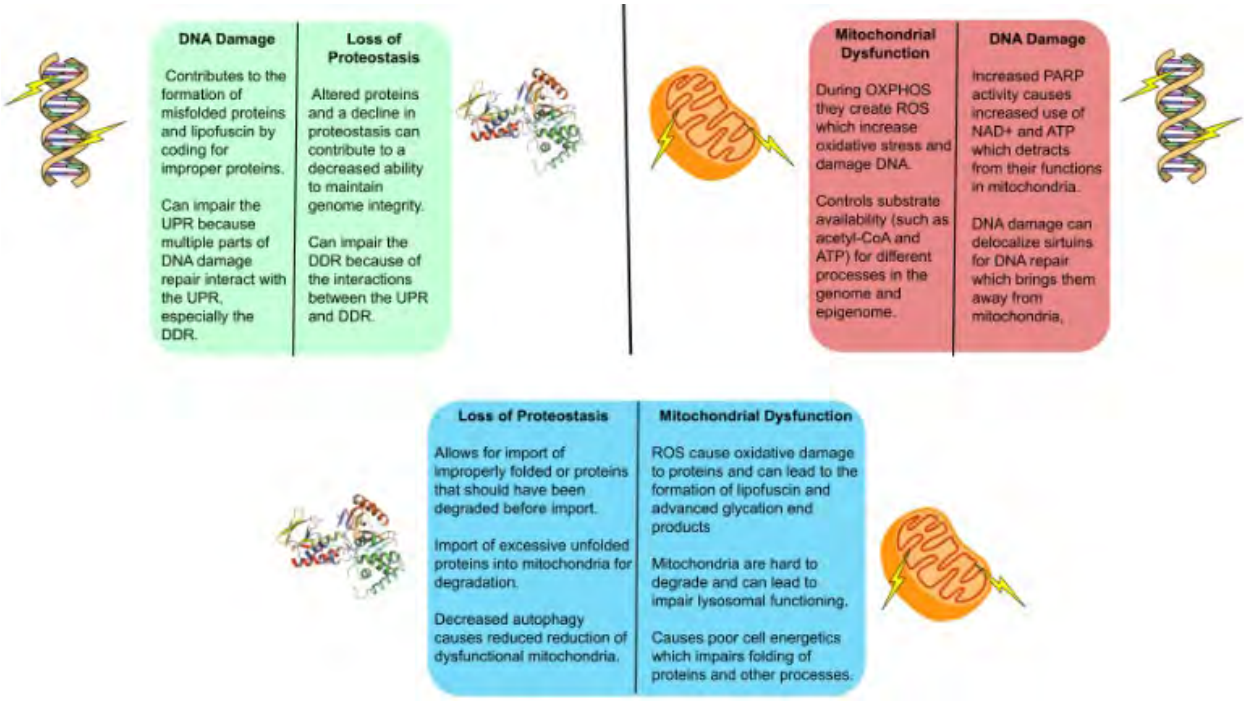


Fig 1: This figure summarizes the different interactions laid out in the text above into their own individual sections. Each box simply shows the more direct influences these causes have on each other, however since aging is so interconnected this should not be taken at face value. Instead, this allows one to see that if any of the primary causes affect any of the other causes, it will cause more damage to every player involved and has many downstream effects. The further downstream effects will be discussed in the next section.

III. The Downstream Causes: Senescent Cells, Stem Cell Exhaustion, and Altered Intercellular Communication.

A. What are these causes?

If the primary causes are seen as the basic causes for aging that are usually deleterious at a cellular level, the downstream causes can be seen as the parts of aging that arise from the primary causes (although not always) and have more system wide effects in how they affect an

organism's aging. This doesn't mean the primary causes don't cause larger systemic effects, mitochondrial dysfunction and loss of proteostasis can both be argued to affect whole organs or body systems, but that the downstream causes are usually more in tissue or system wide effects of aging. Another difference is also that there is usually no context where the primary causes would be inherently beneficial (no cell would be better off with damaged DNA, misfolded proteins, or dysfunctional mitochondria), but senescent cells and inflammation are both used and induced by the body in certain contexts for beneficial effects.

Senescent Cells

Senescent cells are cells that are in a state of cell cycle arrest and have developed morphologic changes. They express changes commonly referred to as the senescence-associated secretory phenotype (SASP). These cells can be formed multiple ways and not all of them are pathological. During embryonic development, cells have been found to be nonproliferative and release p21 and p15, which are all characteristics of SASP [81]. While these cells are removed later in development through apoptosis and immune-mediated clearance, senescent cells seem to play a role in embryonic patterning. They are present in the apical ectodermal ridge (AER) and the neural roof plate, which are both involved in embryonic patterning, and mice deficient in p21 have defects in embryonic senescence, AER maintenance, and patterning [81]. Senescent cells have also been observed to aid in wound healing and suppress cancer formation [82, 83]. During aging, senescent cells usually arise from replicative senescence, which is when a normal cell divides to a point where the telomeres are too short and the cell enters senescence as a preventative mechanism against cancer. These senescent cells are usually removed by the immune system but

can build up as time progresses, resulting in age-related disorders.

Stem Cell Exhaustion

Stem cells are progenitor cells, allowing them to differentiate into multiple other cell types. This trait makes them invaluable in the treatment of age-related diseases and injuries. Mesenchymal stem cells (MSCs) are found at all developmental stages (fetal, young, adult, and aged) and in various tissues, such as bone marrow, synovium, adipose tissue, dental pulp, etc [84, 85]. Mesenchymal stem cells are often used synonymously with multipotent mesenchymal stromal cells, which can be referred to as both MSCs and BM-MSCs, and both of these cells are multipotent stem cells from the bone marrow [85]. These bone marrow derived cells must meet a minimum criteria. They must be plastic-adherent in standard culture conditions, express and have a lack of expression of certain surface molecules, and must be able to differentiate into osteoblasts, adipocytes, and chondroblasts [86]. These cells have also been found to have immunomodulatory properties. MSCs have been reported to regulate the immune response, including T cell and B cell responses [87]. Embryonic Stem Cells (ESCs) are pluripotent cells that can give rise to all somatic cell types in the embryo [88]. Unlike MSCs, ESCs can differentiate into every cell type. These cells are usually produced through in vitro fertilization, and there have been multiple ethical concerns raised about the use of ESCs. Because of this, many scientists are now trying to use other stem cell types for therapies, despite their pluripotent nature.

Altered Intercellular Communication

Intercellular communication involves the communication between cells, which is done through multiple signaling molecules. In aging, the disruption of intercellular communication mainly consists of age-related chronic inflammation, which is

greatly influenced by the immune system. The immune system consists of a complex and very well-regulated network of cells responsible for defending the body against both outside invaders such as viruses and bacteria or against altered components of the host organism, such as cancerous cells or posttranslationally modified macromolecules [89]. The immune system consists of two branches: the innate immune system and the adaptive immune system. The innate immune system is the first line of defense and is meant to rapidly recognize and eliminate pathogens, external threats, danger-associated molecular patterns, and internal threats through the use of specific receptors [90, 91]. The adaptive immune system is a more specific response that develops memory for repeated challenges, which enables qualitatively and quantitatively different responses to reexposures to the same pathogens, allowing for a more efficient response [89]. The adaptive immune system, at least initially, involves fewer cell types than the innate immune system, consisting mainly of T cells and B cells.

As someone ages, multiple age associated changes occur in the immune system, leading to “immunosenescence”. Immunosenescence refers to the age-related decline of immune cells and immune functions, which leads to a higher incidence of infection, cancer, and autoimmune disease in the elderly [92]. There are multiple factors associated with immunosenescence, such as the loss of certain cell types, and one key factor is inflammaging. Inflammaging refers to the chronic, low-grade inflammation that characterizes aging [93]. It develops gradually through the continuous antigenic stimulation in aged subjects which can be provided by pathogens (such as cytomegalovirus, herpes simplex virus-1), cellular and molecular debris arising from

transformations caused by ROS, the Maillard reaction, nitrosylation, or cancer cells [89, 94]. Aside from inflammaging, anti-inflammation has also been proposed. Anti-inflammation represents the changes that counteract inflammaging and may be considered an active phenomenon as inflammation resolves not only through the absence of pro-inflammatory signals but also through activation of specific inhibitory pathways. [95-97].

B. Interactions involving the downstream causes.

1. Stem cells interactions with the primary causes and senescent cells.

Stem cells interact with senescent cells and the primary causes in numerous ways. With mitochondrial dysfunction, decreased cell energetics and increased ROS can damage the cell, leading to stem cell aging and dysfunction. It has been found that older human MSCs have elevated ROS [98]. In mice, the number of hematopoietic stem cells (HSCs) with low ROS levels declines with age in contrast to stem cells with high ROS, and HSCs in low ROS populations have a higher self-renewal potential [99]. Moreover, high cellular ROS concentrations lead to abnormal proliferation, telomere damage, modulation of signaling pathways essential for maintaining HSC quiescence, malignancy, and compromised stem cell self-renewal [100, 101]. The decreased functioning of stem cells with high ROS was found to be restored through treatment with antioxidants, which further shows that ROS damage is a major player in decreased stem cell functioning [99].

In regards to proteostasis, it has been shown that human ESCs have high

proteasome activity, and it has been suggested that proteostasis maintenance in stem cells might have an important role in aging [102]. In mice, HSCs that had a deletion of Atg7 resulted in the loss of normal HSC functions, severe myeloproliferation, and death of the mice within weeks, which suggests the importance of autophagy in stem cells [103]. Additionally, in Atg12 knockout mice, it was found that HSCs had an accumulation of mitochondria and activated metabolic state, driving accelerated myeloid differentiation, impairing HSC self renewal activity, and reducing regenerative potential [104]. These altered metabolic and functional features are also seen in the majority of HSCs in aged mice. Therefore, it seems that autophagy suppresses HSC metabolism and is needed to maintain quiescence and stemness, which becomes increasingly necessary with age to preserve regenerative capacity [104]. This further shows the role of autophagy in stem cell health.

In relation to genetic/epigenetic damage, stem cell exhaustion is also connected with senescent cells, as senescent cells created by genetic and epigenetic damage can cause stem cell exhaustion. It has been shown that DNA damage from oxidative stress activates the expression of cell-cycle inhibitors, leading to the premature senescence and accumulation of senescent cells among HSCs, and this ultimately leads to the loss of stem cell function [105]. This seems to be especially prevalent in stem cells; many stem cells spend long amounts of time in quiescence, which protects them from telomere shortening caused by cell replication [106]. While quiescence protects against this replicative damage, it also makes cells more susceptible to acquiring mutations when DNA damage does occur. This is because when double-strand breaks occur, they are

more likely to be repaired by non-homologous end joining instead of homologous recombination, and non-homologous end joining is more prone to error [50]. Additionally, while quiescence protects against telomere shortening from replication, it does not protect telomeres from all damage. Telomeres have been found to be shortened with age in multiple types of stem cells, despite the presence of telomerase in stem cells [107, 108]. Short telomeres are a hallmark of aging and gradual telomere shortening can also lead to the formation of senescent cells [10].

Finally, links have also been found between epigenetic alterations and stem cells. Aging HSCs show site-specific increases in DNA methylation that are particularly targeted to regions of the genome important for lineage-specific gene expression. Both physiological aging and experimentally enforced proliferation of HSCs led to DNA methylation of genes regulated by Polycomb Repressive Complex 2 [109]. Additionally, it has been found that levels of Histone H4 Lysine 16 Acetylation (AcH4K16) decrease with age in HSCs and inhibition of CDC42 restores AcH4K16 levels to that of young HSCs and reverses phenotypes of HSC aging [110]. Moreover, the expression levels of chromatin modifiers (parts of the SWI-SNF and PRC complexes, HDACs including sirtuins, and DNA methyltransferases) also change with age in stem cells, and these changes may underpin declining stem cell function [50].

2. *Senescent cells interactions with primary causes and altered intercellular communication*

Senescent cells have their own interactions with the primary causes and

another cause of aging, altered intercellular communication. Genetic/epigenetic damage can be seen as one of the main causes for senescent cells. Extensive genetic/epigenetic damage can cause cancer, so cells will become senescent cells to stop the creation of these cancers by stopping cellular division. Using an adenovirus-based system based on tetracycline-controlled expression of the SacI restriction enzyme to introduce DSBs into the mouse liver increases the burden of senescent cells. This triggers elevated levels of some of the aging phenotypes observed in the naturally aged mice, such as increased mitochondrial fusion and apoptotic cells [111].

Epigenetically, it has been found that a distinct heterochromatin structure, where the chromosome is packaged into tightly compact structures called senescence-associated heterochromatin foci (SAHF), accumulates in senescent cells [112, 113]. Studies have also shown that telomeric and subtelomeric regions can have histone modifications found in heterochromatin, and alterations in the epigenetic regulation and DNA methylation at these sites can impact telomere integrity and change telomere length [112, 114, 115]. This is significant because, as stated above, telomere shortening can cause cellular senescence. As the telomeres shorten through repeated DNA replication, eventually the telomeres will reach a critical short length. This impairment will lead to the activation of the p53 or p16INK4a pathway and result in cellular senescence or apoptosis [116]. Additionally, genetic attenuation of the DDR enables reversal of cellular senescence [117, 118]. In mice, activation of the DDR, including formation of DNA damage foci containing activated H2A.X at either uncapped telomeres or persistent DNA strand breaks, is a major trigger of cell senescence, and this supports how the attenuation of the DDR reverses cellular

senescence [119]. The DDR, as stated previously, is also very interconnected with the UPR and proteostasis along with mitochondrial health, which also influence senescent cells in their own rights.

In rat hippocampal cells, proteostasis failure can cause cellular senescence by allowing for protein aggregation, including aggregation of amyloid-β, which is said to be a major contributor to AD. The rat hippocampal cells exhibited a broad range of features that are indicative of senescence, such as SA-β-gal activity, p16 upregulation, lamin B1 loss, the SASP, and stress-resistant phenotypes [120]. Moreover, it was found that downregulation of the mammalian target of rapamycin (mTOR) pathway by rapamycin mitigates the senescence-like phenotypes in the cultured neurons. This suggests that proteostasis is involved in senescence, since mTOR is significantly involved in modulating autophagy and protein synthesis [120]. In human fibroblasts, senescent cells showed multiple aspects of dysfunctional proteostasis. A significant deterioration in the transcriptional activation of the heat shock response was found in the senescent cells, and phosphorylated heat shock factor 1 localization and distribution were impaired in senescent cells [121]. Moreover, alternative splicing regulation was found to be dampened, and decoupling between different UPR branches was found in stressed senescent cells. Similarly, stressed senescent cells had an inability to activate the UPR related transcriptional responses, despite senescent cells showing enhanced translational regulation and endoplasmic reticulum stress sensing [121]. This was accompanied by diminished ATF6 nuclear localization in stressed senescent cells and impaired proteasome functioning. Additionally, senescence has been found to be caused by proteostasis failure in human primary bronchial epithelial cells, human

umbilical vein endothelial cells, and mouse glial cells [120].

With mitochondrial dysfunction, it has been found that inhibition of complex I, II and III in the electron transport chain can induce senescence, and when cells are treated with FCCP to depolarize the mitochondrial membrane potential, this has also been found to induce senescence [25, 122, 123]. These findings reveal that the deterioration of OXPHOS can lead to cellular senescence. In senescent fibroblasts, an increase in mitochondrial mass and abundance of tricarboxylic acid (TCA) cycle metabolites have been observed. Additionally, while there seem to be more mitochondria present in senescent cells, they appear less functional [124]. Furthermore, mitochondria from senescent cells show decreased mitochondrial membrane potential, increased proton leak, and increased generation of ROS. It has been reported as well that these mitochondria have decreased fatty acid oxidation, which can result in increased lipid accumulation [124]. While mitochondrial dysfunction has been found to contribute to the SASP in common senescent cells, it seems it can also cause senescence with a different secretory phenotype: one that lacks the IL-1-dependent inflammatory arm [25, 125]. Finally, as stated previously, mitochondrial dysfunction creates ROS which causes damage to the DNA and the epigenome, leading to cellular senescence.

Senescent cells interact with inflammaging through the SASP and immune system. As immunosenescence takes place and the immune system becomes less effective, there is an increase in senescent T cells along with memory CD4+ and CD8+ T cells. While this is happening, there is also a decrease in the number of naive T cells [126, 127]. These changes could allow for an increase in senescent cell populations. CD8+ T cells target senescent

cells by interacting with NKG2D ligands, and CD4+ T cells are required for proper macrophage dependent elimination of senescent cells [128]. Senescent cells have actually been found to express the non-classical MHC molecule HLA-E to evade immune clearance by specialized CD8+ T cells and natural killer cells, which is released with the SASP [129]. As immune surveillance of senescent cells decreases and they increase in number, the senescent cells secrete pro-inflammatory signals as part of the SASP, which contribute to inflammaging. As inflammaging progresses it may lead to chronic antigen stimulation of T cells and B cells, which can then cause further increases in memory cells and dysfunctional cells. This leads to decreased removal of senescent cells, which results in more pro-inflammatory signals being secreted, creating a positive feedback loop. In accordance with this, it has been found that elimination of senescent cells using a compound based on the increased activity of lysosomal β-galactosidase (a primary characteristic of senescent cells) attenuated low grade local and systemic inflammation and restored physical function [130].

3. *Altered intercellular communication interactions with the primary causes and stem cells*

Altered intercellular communication also has its own interactions with the primary causes and stem cells. Mitochondrial dysfunction heavily contributes to altered intercellular communication by secreting damage associated molecular patterns (DAMPs). DAMPs are signals released by cells if they are experiencing stress, apoptosis, or necrosis. Mitochondria can also secrete

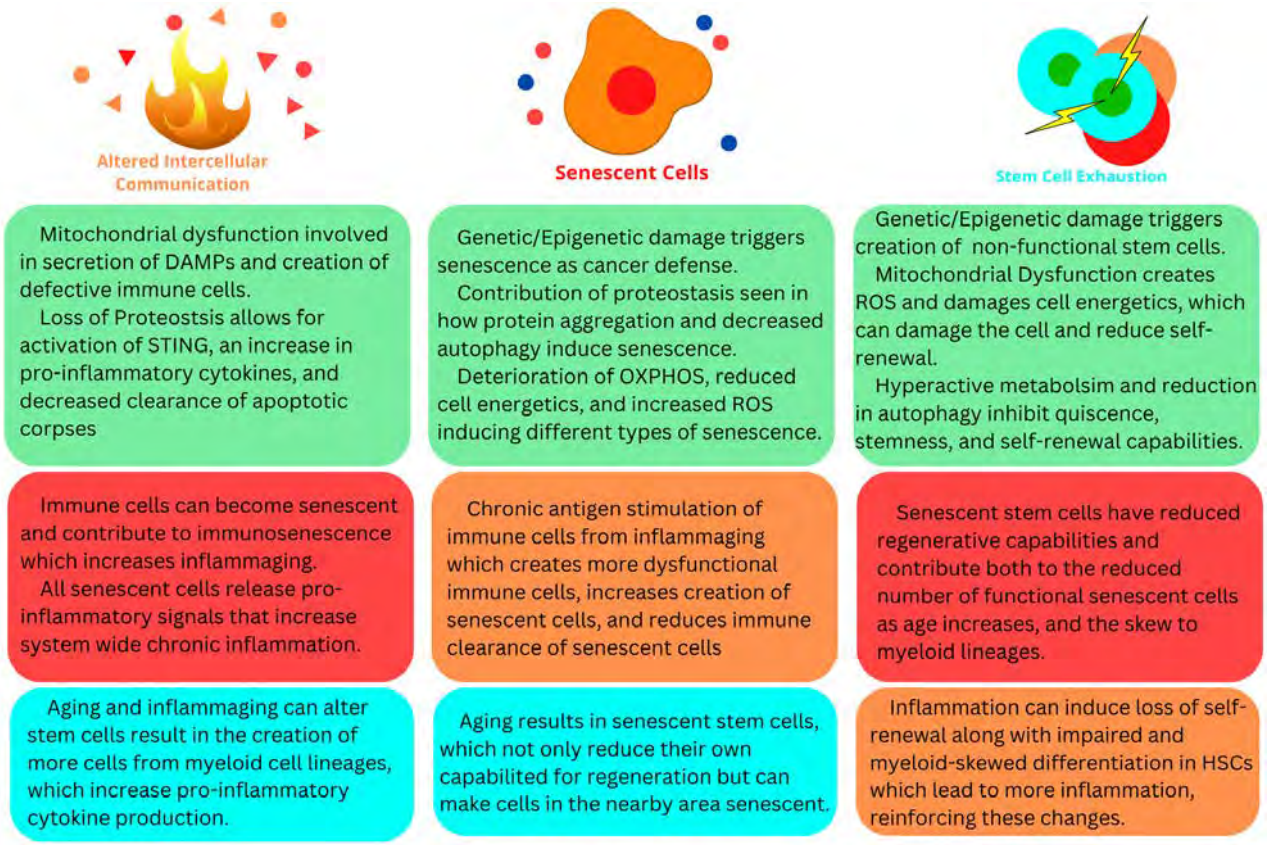
DAMPs, such as cardiolipin, n-formyl peptides, mitochondrial transcription factor A, ATP, mtDNA and ROS [25, 131]. In regards to mtDNA, it was found that plasma levels of mtDNA gradually increase after 50 and correlate with elevated levels of pro-inflammatory cytokines, such as TNF-α, IL-6, RANTES, and IL-1ra. This indicates that mtDNA may promote the production of pro-inflammatory cytokines in aging [131]. Mitochondria also seem to play a role in the loss of immune function. Mice with T cells that were deficient in a mtDNA–stabilizing protein showed multiple age related changes, such as neurological, metabolic, muscular, and cardiovascular impairments, along with defective T cells that initiate both an early inflammatory program and premature senescence [25, 132]. Moreover, metformin has been found to enhance mitochondrial functioning and autophagy, resulting in the alleviation of aging associated inflammation. This supports the impact of mitochondria on inflammaging and hints that proteostasis is also involved in inflammaging [133].

In regards to proteostasis, it has been shown that autophagy protein ATG9A negatively regulates the activation of STING, a transmembrane protein required for efficient activation of type I IFN and pro-inflammatory cytokine production [134]. Additionally, autophagy-deficient cells show increased levels of the adaptor protein p62, which activates the pro-inflammatory transcription factor NF-κB and results in enhanced activity of the stress-responsive transcription factor NRF2 and NRF2-dependent liver injury [134]. However, when properly functioning, autophagy is responsible for the efficient clearance of apoptotic corpses during development and tissue homeostasis which prevents secondary necrosis and the release of DAMPs, limiting inflammation [134]. Autophagy also works with inflammasomes,

however, not much is known about their interactions.

Genetic/epigenetic damage interacts with altered intercellular communication in a multitude of ways as well. Inflammation can result in the release of ROS or reactive nitrogen species from epithelial and immune cells, such as macrophages and neutrophils, which can cause DNA damage [135]. This inflammation can lead to cancers as inflammation is often associated with NFκB activation. NFκB acts antagonistically to p53, a crucial element in DNA damage sensing that is commonly utilized in chemotherapy. Overexpression of NFκB might inhibit p53 functioning because the inhibition of NFκB through overexpression of IκBα in mice embryonic fibroblasts sensitizes cells to chemotherapeutic agents [135]. As shown, the primary causes affect both altered intercellular communication and stem cell exhaustion in significant ways. Both of these factors also influence each other directly. Inflammation can induce a loss of self-renewal and impaired and myeloid-skewed differentiation in HSCs, which resembles many prominent phenotypes of the aging hematopoietic system [136]. While this increases stem cell exhaustion, it also furthers immunosenescence and inflammaging. Pro-inflammatory cytokines are known to be produced by myeloid cell lineages, which can cause a positive feedback loop where HSCs with myeloid-skewed differentiation accumulate in the bone marrow and give rise to more myeloid cells that produce pro-inflammatory cytokines [137]. This advances inflammaging and the myeloid-skewed differentiation of HSCs.

Fig 2: This figure summarizes the different interactions of the downstream causes, similar to figure 1. Under each of the downstream causes are different colored boxes representing what is contributing to that specific cause, with green being the primary causes that were laid out earlier in the text, the red representing senescent cells, the orange representing altered intercellular



communication, and the blue representing stem cells. If one looks under senescent cells they will see how the other factors contribute to the formation and pathology of senescent cells within their respective boxes, and it works similarly for the other downstream causes.

IV. Conclusion

There are many methods of interaction between the various causes of aging, and their complexity can be seen when they are all considered together. For example, if we consider the increased release of pro-inflammatory cytokines from the SASP, we have seen that this directly interacts with stem cells by reducing their regenerative capacity and increasing their myeloid-specific differentiation. At the same time, it heightens inflammaging by increasing the amount of chronic inflammation in an area. This release of pro-inflammatory cytokines also causes chronic antigen stimulation among immune cells, which increases immunosenescence and

causes less immune clearance of senescent cells. This further exacerbates the chronic antigen stimulation and release of pro-inflammatory cytokines. This can also cause cancer or genetic/epigenetic damage by disrupting p53 activity, and the DNA damage causes further creation of senescent cells, while also having its own impacts. The DNA repair disruption consumes NAD+ and uses sirtuins which are needed for proper functioning of the mitochondria. As a result, more mitochondria become dysfunctional, release more ROS, and reduce substrate availability, all of which further increase genetic/epigenetic damage while disrupting repair. This also disrupts proteostasis, since ROS alter proteins which can become aggregated masses and become harder to degrade. Dysfunctional mitochondria also result in reduced cell energetics and reduced substrate availability, decreasing the capacity of the PN to deal with the misfolded proteins. This influences mitophagy which means the dysfunctional mitochondria causing these problems are not removed optimally. This allows them to do more damage, such as inducing senescence, reducing stem cell function, and releasing DAMPs which exacerbate the problem of chronic inflammation.

As shown, the simple release of pro-inflammatory cytokines from senescent cells is interconnected with multiple other causes of aging through their own direct interactions. Usually, some extra pro-inflammatory signaling would not be a problem for the body, but having it occur for years on end along with all of the other factors in aging starts to paint a picture of why aging is difficult to combat. In connecting the multifaceted interactions between these factors, a much more complex and complete picture of biological aging is revealed. However, it should be noted that many interactions are likely still undiscovered, and some of the interactions

that have been discovered are still being investigated. The discussed interactions do portray the interconnectedness and complexity between the causes of aging, but further research is still required to fully understand aging.

Conflict Of Interest

The author declares no conflict of interest.

Funding

This paper was not funded.

Author Contributions

The single author is the only one that wrote this paper.

Ethics Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] Hara, Y., McKeehan, N., & Fillit, H. M. (2018). Translating the biology of aging into novel therapeutics for Alzheimer disease. Neurology, 92(2), 84–93. https://doi.org/10.1212/wnl.0000000000006745

[2] De Grey, A. D. N. J, & Rae, M. (2007). Ending aging: the rejuvenation breakthroughs that could reverse human aging in our lifetime. St. Martin's Griffin.

[3] López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The Hallmarks of Aging. Cell, 153(6), 1194–1217. https://doi.org/10.1016/j.cell.2013.05.039

[4] Tiwari, V., & Wilson, D. M. (2019). DNA Damage and Associated DNA Repair Defects in Disease and Premature Aging. The American Journal of Human Genetics, 105(2), 237–257. https://doi.org/10.1016/j.ajhg.2019.06.005

[5] Li, G.-M. (2007). Mechanisms and functions of DNA mismatch repair. Cell Research, 18(1), 85–98. https://doi.org/10.1038/cr.2007.115

[6] Best, B. P. (2009). Nuclear DNA Damage as a Direct Cause of Aging. Rejuvenation Research, 12(3), 199–208. https://doi.org/10.1089/rej.2009.0847

[7] Hasty, P., Campisi, J., Hoeijmakers, J., Steeg, H. van, & Vijg, J. (2003). Aging and Genome Maintenance: Lessons from the Mouse? Science, 299(5611), 1355–1359. https://doi.org/10.1126/science.1079161

[8] Ljungman, M., & Lane, D. P. (2004). Transcription — guarding the genome by sensing DNA damage. Nature Reviews Cancer, 4(9), 727–737. https://doi.org/10.1038/nrc1435

[9] Griffith, J. D., Comeau, L., Rosenfield, S., Stansel, R. M., Bianchi, A., Moss, H., & de Lange, T. (1999). Mammalian telomeres end in a large duplex loop. Cell, 97(4), 503–514. https://doi.org/10.1016/s0092-8674(00)80760-6

[10] Shin, J.-S., Hong, A., Solomon, M. J., & Lee, C. S. (2006). The role of telomeres and telomerase in the pathology of human cancer and aging. Pathology, 38(2), 103–113. https://doi.org/10.1080/00313020600580468

[11] Zhu, Y., Liu, X., Ding, X., Wang, F., & Geng, X. (2018). Telomere and its role in the aging pathways: telomere shortening, cell senescence and mitochondria dysfunction. Biogerontology, 20(1), 1–16. https://doi.org/10.1007/s10522-018-9769-1

[12] Pal, S., & Tyler, J. K. (2016). Epigenetics and aging. Science Advances, 2(7). https://doi.org/10.1126/sciadv.1600584

[13] Braga, D. L., Mousovich-Neto, F., Tonon-Da-Silva, G., Salgueiro, W. G., & Mori, M. A. (2020). Epigenetic changes during ageing and their underlying mechanisms. Biogerontology, 21(4), 423–443. https://doi.org/10.1007/s10522-020-09874-y

[14] Powers, E. T., Morimoto, R. I., Dillin, A., Kelly, J. W., & Balch, W. E. (2009). Biological and Chemical Approaches to Diseases of Proteostasis Deficiency. Annual Review of Biochemistry, 78(1), 959–991. https://doi.org/10.1146/annurev.biochem.052308.114844

[15] Labbadia, J., & Morimoto, R. I. (2014). Proteostasis and longevity: when does aging really begin? F1000Prime Reports, 6. https://doi.org/10.12703/p6-07

[16] Balch, W. E., Morimoto, R. I., Dillin, A., & Kelly, J. W. (2008). Adapting Proteostasis for Disease Intervention. Science, 319(5865), 916–919. https://doi.org/10.1126/science.1141448

[17] Klaips, C. L., Jayaraj, G. G., & Hartl, F. U. (2017). Pathways of cellular proteostasis in aging and disease. Journal of Cell Biology, 217(1), 51–63. https://doi.org/10.1083/jcb.201709072

[18] Hartl, F. U. (1996). Molecular chaperones in cellular protein folding. Nature, 381(6583), 571–580. https://doi.org/10.1038/381571a0

[19] Kaushik, S., & Cuervo, A. M. (2015). Proteostasis and aging. Nature Medicine, 21(12), 1406–1415. https://doi.org/10.1038/nm.4001

[20] Ding, W.-X., & Yin, X.-M. (2008). Sorting, recognition and activation of the misfolded protein degradation pathways through macroautophagy and the proteasome. Autophagy, 4(2), 141–150. https://doi.org/10.4161/auto.5190

[21] Pohl, C., & Dikic, I. (2019). Cellular quality control by the ubiquitin-proteasome system and autophagy. Science, 366(6467), 818–822. https://doi.org/10.1126/science.aax3769

[22] Kauppila, T. E., Kauppila, J. H., & Larsson, N.-G. (2017). Mammalian Mitochondria and Aging: An Update. Cell Metabolism, 25(1), 57–71. https://doi.org/10.1016/j.cmet.2016.09.017

[23] Houtkooper, R. H., Argmann, C., Houten, S. M., Cantó, C., Jeninga, E. H., Andreux, P. A., … Auwerx, J. (2011). The metabolic footprint of aging in mice. Scientific Reports, 1(1). https://doi.org/10.1038/srep00134

[24] Lesnefsky, E. J., & Hoppel, C. L. (2006). Oxidative phosphorylation and aging. Ageing Research Reviews, 5(4), 402–433. https://doi.org/10.1016/j.arr.2006.04.001

[25] van der Rijt, S., Molenaars, M., McIntyre, R. L., Janssens, G. E., & Houtkooper, R. H. (2020). Integrating the hallmarks of aging throughout the tree of life: A focus on mitochondrial dysfunction. Frontiers in Cell and Developmental Biology, 8. https://doi.org/10.3389/fcell.2020.594416

[26] Hipp, M. S., Park, S.-H., & Hartl, F. U. (2014). Proteostasis impairment in protein-misfolding and -aggregation diseases. Trends in Cell Biology, 24(9), 506–514. https://doi.org/10.1016/j.tcb.2014.05.003

[27] Moreno-García, A., Kun, A., Calero, O., Medina, M., & Calero, M. (2018). An Overview of the Role of Lipofuscin in Age-Related Neurodegeneration. Frontiers in Neuroscience, 12. https://doi.org/10.3389/fnins.2018.00464

[28] Press, M., Jung, T., König, J., Grune, T., & Höhn, A. (2019). Protein aggregates and proteostasis in aging: Amylin and β-cell function. Mechanisms of Ageing and Development, 177, 46–54. https://doi.org/10.1016/j.mad.2018.03.010

[29] Gavande, N. S., Vandervere-Carozza, P. S., Hinshaw, H. D., Jalal, S. I., Sears, C. R., Pawelczak, K. S., & Turchi, J. J. (2016). DNA repair targeted therapy: The past or future of cancer treatment? Pharmacology & Therapeutics, 160, 65-83. doi:10.1016/j.pharmthera.2016.02.003

[30] Shay, J. W., & Keith, W. N. (2008). Targeting telomerase for cancer therapeutics. British Journal of Cancer, 98(4), 677-683. doi:10.1038/sj.bjc.6604209

[31] Barbosa, M. C., Grosso, R. A., & Fader, C. M. (2019). Hallmarks of Aging: An Autophagic Perspective. Frontiers in Endocrinology, 9. https://doi.org/10.3389/fendo.2018.00790

[32] Khalil, H., Tazi, M., Caution, K., Ahmed, A., Kanneganti, A., Assani, K., … Amer, A. O. (2016). Aging is associated with hypermethylation of autophagy genes in macrophages. Epigenetics, 11(5), 381–388. https://doi.org/10.1080/15592294.2016.1144007

[33] González-Quiroz, M., Blondel, A., Sagredo, A., Hetz, C., Chevet, E., & Pedeux, R. (2020). When Endoplasmic Reticulum Proteostasis Meets the DNA Damage Response. Trends in Cell Biology, 30(11), 881–891. https://doi.org/10.1016/j.tcb.2020.09.002

[34] Zheng, P., Chen, Q., Tian, X., Qian, N., Chai, P., Liu, B., … Chen, J. (2018). DNA damage triggers tubular endoplasmic reticulum extension to promote apoptosis by facilitating ER-mitochondria signaling. Cell Research, 28(8), 833–854. https://doi.org/10.1038/s41422-018-0065-z

[35] Bratic, I., & Trifunovic, A. (2010). Mitochondrial energy metabolism and ageing. Biochimica Et Biophysica Acta (BBA) - Bioenergetics, 1797(6-7), 961–967. https://doi.org/10.1016/j.bbabio.2010.01.004

[36] Cedikova, M., Pitule, P., Kripnerova, M., Markova, M., & Kuncova, J. (2016). Multiple Roles of Mitochondria in Aging Processes. Physiological Research. https://doi.org/10.33549/physiolres.933538

[37] Dai, D.-F., Chiao, Y., Marcinek, D. J., Szeto, H. H., & Rabinovitch, P. S. (2014). Mitochondrial oxidative stress in aging and healthspan. Longevity & Healthspan, 3(1), 6. https://doi.org/10.1186/2046-2395-3-6

[38] Liu, Y., Fiskum, G., & Schubert, D. (2002). Generation of reactive oxygen species by the mitochondrial electron transport chain. Journal of Neurochemistry, 80(5), 780–787. https://doi.org/10.1046/j.0022-3042.2002.00744.x

[39] Marnett, L. J. (1999). Lipid peroxidation—DNA damage by malondialdehyde. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 424(1-2), 83–95. https://doi.org/10.1016/s0027-5107(99)00010-x

[40] Fakouri, N. B., Hou, Y., Demarest, T. G., Christiansen, L. S., Okur, M. N., Mohanty, J. G., … Bohr, V. A. (2018). Toward understanding genomic instability, mitochondrial dysfunction and aging. The FEBS Journal, 286(6), 1058–1073. https://doi.org/10.1111/febs.14663

[41] Éthier, C., Tardif, M., Arul, L., & Poirier, G. G. (2012). PARP-1 Modulation of mTOR Signaling in Response to a DNA Alkylating Agent. PLoS ONE, 7(10). https://doi.org/10.1371/journal.pone.0047978

[42] Ha, H. C., & Snyder, S. H. (1999). Poly(ADP-ribose) polymerase is a mediator of necrotic cell death by ATP depletion. Proceedings of the National Academy of Sciences, 96(24), 13978–13982. https://doi.org/10.1073/pnas.96.24.13978

[43] Bai, P., Cantó, C., Oudart, H., Brunyánszki, A., Cen, Y., Thomas, C., … Auwerx, J. (2011). PARP-1 Inhibition Increases Mitochondrial Metabolism through SIRT1 Activation. Cell Metabolism, 13(4), 461–468. https://doi.org/10.1016/j.cmet.2011.03.004

[44] Mouchiroud, L., Houtkooper, R. H., Moullan, N., Katsyuba, E., Ryu, D., Cantó, C., … Auwerx, J. (2013). The NAD+/Sirtuin Pathway Modulates Longevity through Activation of Mitochondrial UPR and FOXO Signaling. Cell, 154(2), 430–441. https://doi.org/10.1016/j.cell.2013.06.016

[45] Pirinen, E., Cantó, C., Jo, Y. S., Morato, L., Zhang, H., Menzies, K. J., … Auwerx, J. (2014). Pharmacological Inhibition of Poly(ADP-Ribose) Polymerases Improves Fitness and Mitochondrial Function in Skeletal Muscle. Cell Metabolism, 19(6), 1034–1041. https://doi.org/10.1016/j.cmet.2014.04.002

[46] Balan, V., Miller, G. S., Kaplun, L., Balan, K., Chong, Z.-Z., Li, F., … Tzivion, G. (2008). Life Span Extension and Neuronal Cell Protection by Drosophila Nicotinamidase. Journal of Biological Chemistry, 283(41), 27810–27819. https://doi.org/10.1074/jbc.m804681200

[47] Fang, E. F., Lautrup, S., Hou, Y., Demarest, T. G., Croteau, D. L., Mattson, M. P., & Bohr, V. A. (2017). NAD + in Aging: Molecular Mechanisms and Translational Implications. Trends in Molecular Medicine, 23(10), 899–916. https://doi.org/10.1016/j.molmed.2017.08.001

[48] Fang, E. F., Kassahun, H., Croteau, D. L., Scheibye-Knudsen, M., Marosi, K., Lu, H., … Bohr, V. A. (2016). NAD + Replenishment Improves Lifespan and Healthspan in Ataxia Telangiectasia Models via Mitophagy and DNA Repair. Cell Metabolism, 24(4), 566–581. https://doi.org/10.1016/j.cmet.2016.09.004

[49] Zhang, H., Ryu, D., Wu, Y., Gariani, K., Wang, X., Luan, P., … Auwerx, J. (2016). NAD+ repletion improves mitochondrial and stem cell function and enhances life span in mice. Science, 352(6292), 1436–1443. https://doi.org/10.1126/science.aaf2693

[50] Schultz, M. B., & Sinclair, D. A. (2016). Why NAD + Declines during Aging: It's Destroyed. Cell Metabolism, 23(6), 965–966. https://doi.org/10.1016/j.cmet.2016.05.022

[51] Houtkooper, R. H., Pirinen, E., & Auwerx, J. (2012). Sirtuins as regulators of metabolism and healthspan. Nature Reviews Molecular Cell Biology, 13(4), 225–238. https://doi.org/10.1038/nrm3293

[52] van de Ven, R. A. H., Santos, D., & Haigis, M. C. (2017). Mitochondrial Sirtuins and Molecular Mechanisms of Aging. Trends in Molecular Medicine, 23(4), 320–331. https://doi.org/10.1016/j.molmed.2017.02.005

[53] Lagunas-Rangel, F. A. (2019). Current role of mammalian sirtuins in DNA repair. DNA Repair, 80, 85–92. https://doi.org/10.1016/j.dnarep.2019.06.009

[54] Sengupta, A., & Haldar, D. (2018). Human sirtuin 3 (SIRT3) deacetylates histone H3 lysine 56 to promote nonhomologous end joining repair. DNA Repair, 61, 1–16. https://doi.org/10.1016/j.dnarep.2017.11.003

[55] Kadam, A., Jubin, T., Roychowdhury, R., & Begum, R. (2020). Role of PARP-1 in mitochondrial homeostasis. Biochimica Et Biophysica Acta (BBA) - General Subjects, 1864(10), 129669. https://doi.org/10.1016/j.bbagen.2020.129669

[56] Yoon, S. P., & Kim, J. (2016). Poly(ADP-ribose) polymerase 1 contributes to oxidative stress through downregulation of sirtuin 3 during cisplatin nephrotoxicity. Anatomy & Cell Biology, 49(3), 165. https://doi.org/10.5115/acb.2016.49.3.165

[57] Ahn, B.-H., Kim, H.-S., Song, S., Lee, I. H., Liu, J., Vassilopoulos, A., … Finkel, T. (2008). A role for the mitochondrial deacetylase Sirt3 in regulating energy homeostasis. Proceedings of the National Academy of Sciences, 105(38), 14447–14452. https://doi.org/10.1073/pnas.0803790105

[58] Wątroba, M., Dudek, I., Skoda, M., Stangret, A., Rzodkiewicz, P., & Szukiewicz, D. (2017). Sirtuins, epigenetics and longevity. Ageing Research Reviews, 40, 11–19. https://doi.org/10.1016/j.arr.2017.08.001

[59] Matilainen, O., Quirós, P. M., & Auwerx, J. (2017). Mitochondria and Epigenetics – Crosstalk in Homeostasis and Stress. Trends in Cell Biology, 27(6), 453–463. https://doi.org/10.1016/j.tcb.2017.02.004

[60] Aon, M. A., Cortassa, S., Juhaszova, M., & Sollott, S. J. (2016). Mitochondrial health, the epigenome and healthspan. Clinical Science, 130(15), 1285–1305. https://doi.org/10.1042/cs20160002

[61] Teperino, R., Schoonjans, K., & Auwerx, J. (2010). Histone Methyl Transferases and Demethylases; Can They Link Metabolism and Transcription? Cell Metabolism, 12(4), 321–327. https://doi.org/10.1016/j.cmet.2010.09.004

[62] Greer, E. L., Blanco, M. A., Gu, L., Sendinc, E., Liu, J., Aristizábal-Corrales, D., … Shi, Y. (2015). DNA Methylation on N6-Adenine in C. elegans. Cell, 161(4), 868–878. https://doi.org/10.1016/j.cell.2015.04.005

[63] Wu, T. P., Wang, T., Seetin, M. G., Lai, Y., Zhu, S., Lin, K., … Xiao, A. Z. (2016). DNA methylation on N6-adenine in mammalian embryonic stem cells. Nature, 532(7599), 329–333. https://doi.org/10.1038/nature17640

[64] O'hagan, H. M. (2013). Chromatin modifications during repair of environmental exposure-induced DNA damage: A potential mechanism for stable epigenetic alterations.

[65] Smeenk, G., & Attikum, H. V. (2013). The Chromatin Response to DNA Breaks: Leaving a Mark on Genome Integrity. Annual Review of Biochemistry, 82(1), 55–80. https://doi.org/10.1146/annurev-biochem-061809-174504

[66] Zglinicki, T. V. (2002). Oxidative stress shortens telomeres. Trends in Biochemical Sciences, 27(7), 339–344. https://doi.org/10.1016/s0968-0004(02)02110-2

[67] Moehle, E. A., Shen, K., & Dillin, A. (2019). Mitochondrial proteostasis in the context of cellular and organismal health and aging. Journal of Biological Chemistry, 294(14), 5396–5407. https://doi.org/10.1074/jbc.tm117.000893

[68] Margineantu, D. H., Emerson, C. B., Diaz, D., & Hockenbery, D. M. (2007). Hsp90 Inhibition Decreases Mitochondrial Protein Turnover. PLoS ONE, 2(10). https://doi.org/10.1371/journal.pone.0001066

[69] Quiles, J. M., & Gustafsson, Å. B. (2020). Mitochondrial Quality Control and Cellular Proteostasis: Two Sides of the Same Coin. Frontiers in Physiology, 11. https://doi.org/10.3389/fphys.2020.00515

[70] Lavie, J., De Belvalet, H., Sonon, S., Ion, A. M., Dumon, E., Melser, S., … Bénard, G. (2018). Ubiquitin-Dependent Degradation of Mitochondrial Proteins Regulates Energy Metabolism. Cell Reports, 23(10), 2852–2863. https://doi.org/10.1016/j.celrep.2018.05.013

[71] Rawat, S., Anusha, V., Jha, M., Sreedurgalakshmi, K., & Raychaudhuri, S. (2019). Aggregation of Respiratory Complex Subunits Marks the Onset of Proteotoxicity in Proteasome Inhibited Cells. Journal of Molecular Biology, 431(5), 996–1015. https://doi.org/10.1016/j.jmb.2019.01.022

[72] Karbowski, M., & Youle, R. J. (2011). Regulating mitochondrial outer membrane proteins by ubiquitination and proteasomal degradation. Current Opinion in Cell Biology, 23(4), 476–482. https://doi.org/10.1016/j.ceb.2011.05.007

[73] Ruan, L., Wang, Y., Zhang, X., Tomaszewski, A., McNamara, J. T., & Li, R. (2020). Mitochondria-associated proteostasis. Annual Review of Biophysics, 49(1), 41–67. https://doi.org/10.1146/annurev-biophys-121219-081604

[74] Ruan, L., Zhou, C., Jin, E., Kucharavy, A., Zhang, Y., Wen, Z., … Li, R. (2017). Cytosolic proteostasis through importing of misfolded proteins into mitochondria. Nature, 543(7645), 443–446. https://doi.org/10.1038/nature21695

[75] Hansson Petersen, C. A., Alikhani, N., Behbahani, H., Wiehager, B., Pavlov, P. F., Alafuzoff, I., … Ankarcrona, M. (2008). The amyloid β-peptide is imported into mitochondria via the TOM import machinery and localized to mitochondrial cristae. Proceedings of the National Academy of Sciences, 105(35), 13145–13150. https://doi.org/10.1073/pnas.0806192105

[76] Sorrentino, V., Romani, M., Mouchiroud, L., Beck, J. S., Zhang, H., D'Amico, D., … Auwerx, J. (2017). Enhancing mitochondrial proteostasis reduces amyloid-β proteotoxicity. Nature, 552(7684), 187–193. https://doi.org/10.1038/nature25143

[77] Cenini, G., Rüb, C., Bruderek, M., & Voos, W. (2016). Amyloid β-peptides interfere with mitochondrial preprotein import competence by a coaggregation process. Molecular Biology of the Cell, 27(21), 3257–3272. https://doi.org/10.1091/mbc.e16-05-0313

[78] Youle, R. J., & Narendra, D. P. (2011). Mechanisms of mitophagy. Nature Reviews Molecular Cell Biology, 12(1), 9–14. https://doi.org/10.1038/nrm3028

[79] Ma, Y., & Li, J. (2015). Metabolic Shifts during Aging and Pathology. Comprehensive Physiology, 667–686. https://doi.org/10.1002/cphy.c140041

[80] Chaudhuri, J., Bains, Y., Guha, S., Kahn, A., Hall, D., Bose, N., … Kapahi, P. (2018). The Role of Advanced Glycation End Products in Aging and Metabolic Diseases: Bridging Association and Causality. Cell Metabolism, 28(3), 337–352. https://doi.org/10.1016/j.cmet.2018.08.014

[81] Storer, M., Mas, A., Robert-Moreno, A., Pecoraro, M., Ortells, M. C., Di Giacomo, V., … Keyes, W. M. (2013). Senescence Is a Developmental Mechanism that Contributes to Embryonic Growth and Patterning. Cell, 155(5), 1119–1130. https://doi.org/10.1016/j.cell.2013.10.041

[82] Demaria, M., Ohtani, N., Youssef, S. A., Rodier, F., Toussaint, W., Mitchell, J. R., … Campisi, J. (2014). An Essential Role for Senescent Cells in Optimal Wound Healing through Secretion of PDGF-AA. Developmental Cell, 31(6), 722–733. https://doi.org/10.1016/j.devcel.2014.11.012

[83] He, S., & Sharpless, N. E. (2017). Senescence in Health and Disease. Cell, 169(6), 1000–1011. https://doi.org/10.1016/j.cell.2017.05.015

[84] Andrzejewska, A., Lukomska, B., & Janowski, M. (2019). Concise Review: Mesenchymal Stem Cells: From Roots to Boost. Stem Cells, 37(7), 855–864. https://doi.org/10.1002/stem.3016

[85] Bianco, P., Robey, P. G., & Simmons, P. J. (2008). Mesenchymal Stem Cells: Revisiting History, Concepts, and Assays. Cell Stem Cell, 2(4), 313–319. https://doi.org/10.1016/j.stem.2008.03.002

[86] Dominici, M., Blanc, K. L., Mueller, I., Slaper-Cortenbach, I., Marini, F., Krause, D., … Horwitz, E. (2006). Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. Cytotherapy, 8(4), 315–317. https://doi.org/10.1080/14653240600855905

[87] Gao, F., Chiu, S. M., Motan, D. A. L., Zhang, Z., Chen, L., Ji, H.-L., … Lian, Q. (2016). Mesenchymal stem cells and immunomodulation: current status and future prospects. Cell Death & Disease, 7(1). https://doi.org/10.1038/cddis.2015.327

[88] Vazin, T., & Freed, W. J. (2010). Human embryonic stem cells: Derivation, culture, and differentiation: A review. Restorative Neurology and Neuroscience, 28(4), 589–603. https://doi.org/10.3233/rnn-2010-0543

[89] Fulop, T., Witkowski, J. M., Pawelec, G., Alan, C., & Larbi, A. (2014). On the Immunological Theory of Aging. Aging Interdisciplinary Topics in Gerontology, 163–176. https://doi.org/10.1159/000358904

[90] Fulop, T., Larbi, A., Dupuis, G., Page, A. L., Frost, E. H., Cohen, A. A., … Franceschi, C. (2018). Immunosenescence and Inflamm-Aging As Two Sides of the Same Coin: Friends or Foes? Frontiers in Immunology, 8. https://doi.org/10.3389/fimmu.2017.01960

[91] Rivera, A., Siracusa, M. C., Yap, G. S., & Gause, W. C. (2016). Innate cell communication kick-starts pathogen-specific immunity. Nature Immunology, 17(4), 356–363. https://doi.org/10.1038/ni.3375

[92] Stahl, E. C., & Brown, B. N. (2015). Cell Therapy Strategies to Combat Immunosenescence. Organogenesis, 11(4), 159–172. https://doi.org/10.1080/15476278.2015.1120046

[93] Franceschi, C., Zaikin, A., Gordleeva, S., Ivanchenko, M., Bonifazi, F., Storci, G., & Bonafè, M. (2018). Inflammaging 2018: An update and a model. Seminars in Immunology, 40, 1–5. https://doi.org/10.1016/j.smim.2018.10.008

[94] Solana, R., Tarazona, R., Aiello, A. E., Akbar, A. N., Appay, V., Beswick, M., … Pawelec, G. (2012). CMV and Immunosenescence: from basics to clinics. Immunity & Ageing, 9(1). https://doi.org/10.1186/1742-4933-9-23

[95] Giunta, S. (2006). Is inflammaging an auto[innate]immunity subclinical syndrome? Immunity & Ageing, 3(1). https://doi.org/10.1186/1742-4933-3-12

[96] Xia, S., Zhang, X., Zheng, S., Khanabdali, R., Kalionis, B., Wu, J., … Tai, X. (2016). An Update on Inflamm-Aging: Mechanisms, Prevention, and Treatment. Journal of Immunology Research, 2016, 1–12. https://doi.org/10.1155/2016/8426874

[97] Franceschi, C., Capri, M., Monti, D., Giunta, S., Olivieri, F., Sevini, F., … Salvioli, S. (2007). Inflammaging and anti-inflammaging: A systemic perspective on aging and longevity emerged from studies in humans. Mechanisms of Ageing and Development, 128(1), 92–105. https://doi.org/10.1016/j.mad.2006.11.016

[98] Stolzing, A., Jones, E., Mcgonagle, D., & Scutt, A. (2008). Age-related changes in human bone marrow-derived mesenchymal stem cells: Consequences for cell therapies. Mechanisms of Ageing and Development, 129(3), 163–173. https://doi.org/10.1016/j.mad.2007.12.002

[99] Jang, Y.-Y., & Sharkis, S. J. (2007). A low level of reactive oxygen species selects for primitive hematopoietic stem cells that may reside in the low-oxygenic niche. Blood, 110(8), 3056–3063. https://doi.org/10.1182/blood-2007-05-087759

[100] Sahin, E., & Depinho, R. A. (2010). Linking functional decline of telomeres, mitochondria and stem cells during ageing. Nature, 464(7288), 520–528. https://doi.org/10.1038/nature08982

[101] Oh, J., Lee, Y. D., & Wagers, A. J. (2014). Stem cell aging: mechanisms, regulators and therapeutic opportunities. Nature Medicine, 20(8), 870–880. https://doi.org/10.1038/nm.3651

[102] Vilchez, D., Simic, M. S., & Dillin, A. (2014). Proteostasis and aging of stem cells. Trends in Cell Biology, 24(3), 161–170. https://doi.org/10.1016/j.tcb.2013.09.002

[103] Mortensen, M., Soilleux, E. J., Djordjevic, G., Tripp, R., Lutteropp, M., Sadighi-Akha, E., … Simon, A. K. (2011). The autophagy protein Atg7 is essential for hematopoietic stem cell maintenance. Journal of Experimental Medicine, 208(3), 455–467. https://doi.org/10.1084/jem.20101145

[104] Ho, T. T., Warr, M. R., Adelman, E. R., Lansinger, O. M., Flach, J., Verovskaya, E. V., … Passegué, E. (2017). Autophagy maintains the metabolism and function of young and old (hematopoietic) stem cells. Nature, 543(7644), 205–210. https://doi.org/10.1038/nature21388

[105] Yahata, T., Takanashi, T., Muguruma, Y., Ibrahim, A. A., Matsuzawa, H., Uno, T., … Ando, K. (2011). Accumulation of oxidative DNA damage restricts the self-renewal capacity of human hematopoietic stem cells. Blood, 118(11), 2941–2950. https://doi.org/10.1182/blood-2011-01-330050

[106] Tomasetti, C., & Vogelstein, B. (2015). Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science, 347(6217), 78–81. https://doi.org/10.1126/science.1260825

[107] Bonab, M., Alimoghaddam, K., Talebian, F., Ghaffari, S., Ghavamzadeh, A., & Nikbin, B. (2006). Aging of mesenchymal stem cell in vitro. BMC Cell Biology, 7(1), 14. https://doi.org/10.1186/1471-2121-7-14

[108] Flores, I., Canela, A., Vera, E., Tejera, A., Cotsarelis, G., & Blasco, M. A. (2008). The longest telomeres: a general signature of adult stem cell compartments. Genes & Development, 22(5), 654–667. https://doi.org/10.1101/gad.451008

[109] Beerman, I., Bock, C., Garrison, B. S., Smith, Z. D., Gu, H., Meissner, A., & Rossi, D. J. (2013). Proliferation-Dependent Alterations of the DNA Methylation Landscape Underlie Hematopoietic Stem Cell Aging. Cell Stem Cell, 12(4), 413–425. https://doi.org/10.1016/j.stem.2013.01.017

[110] Florian, M. C., Dörr, K., Niebel, A., Daria, D., Schrezenmeier, H., Rojewski, M., … Geiger, H. (2012). Cdc42 Activity Regulates Hematopoietic Stem Cell Aging and Rejuvenation. Cell Stem Cell, 10(5), 520–530. https://doi.org/10.1016/j.stem.2012.04.007

[111] White, R. R., Milholland, B., Bruin, A. D., Curran, S., Laberge, R.-M., Steeg, H. V., … Vijg, J. (2015). Controlled induction of DNA double-strand breaks in the mouse liver induces features of tissue ageing. Nature Communications, 6(1). https://doi.org/10.1038/ncomms7790

[112] Muñoz-Najar, U., & Sedivy, J. M. (2011). Epigenetic Control of Aging. Antioxidants & Redox Signaling, 14(2), 241–259. https://doi.org/10.1089/ars.2010.3250

[113] Narita, M., Nuñez, S., Heard, E., Narita, M., Lin, A. W., Hearn, S. A., … Lowe, S. W. (2003). Rb-Mediated Heterochromatin Formation and Silencing of E2F Target Genes during Cellular Senescence. Cell, 113(6), 703–716. https://doi.org/10.1016/s0092-8674(03)00401-x

[114] García-Cao, M., O'sullivan, R., Peters, A. H. F. M., Jenuwein, T., & Blasco, M. A. (2003). Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. Nature Genetics, 36(1), 94–99. https://doi.org/10.1038/ng1278

[115] Gonzalo, S., Jaco, I., Fraga, M. F., Chen, T., Li, E., Esteller, M., & Blasco, M. A. (2006). DNA methyltransferases control telomere length and telomere recombination in mammalian cells. Nature Cell Biology, 8(4), 416–424. https://doi.org/10.1038/ncb1386

[116] Blasco, M. A. (2005). Telomeres and human disease: ageing, cancer and beyond. Nature Reviews Genetics, 6(8), 611–622. https://doi.org/10.1038/nrg1656

[117] Niedernhofer, L. J., Gurkar, A. U., Wang, Y., Vijg, J., Hoeijmakers, J. H., & Robbins, P. D. (2018). Nuclear Genomic Instability and Aging. Annual Review of Biochemistry, 87(1), 295–322. https://doi.org/10.1146/annurev-biochem-062917-012239

[118] Beausejour, C. M. (2003). Reversal of human cellular senescence: roles of the p53 and p16 pathways. The EMBO Journal, 22(16), 4212–4222. https://doi.org/10.1093/emboj/cdg417

[119] Wang, C., Jurk, D., Maddick, M., Nelson, G., Martin-Ruiz, C., & Zglinicki, T. V. (2009). DNA damage response and cellular senescence in tissues of aging mice. Aging Cell, 8(3), 311–323. https://doi.org/10.1111/j.1474-9726.2009.00481.x

[120] Ishikawa, S., & Ishikawa, F. (2019). Proteostasis failure and cellular senescence in long-term cultured postmitotic rat neurons. Aging Cell, 19(1). https://doi.org/10.1111/acel.13071

[121] Sabath, N., Levy-Adam, F., Younis, A., Rozales, K., Meller, A., Hadar, S., … Shalgi, R. (2020). Cellular proteostasis decline in human senescence. Proceedings of the National Academy of Sciences, 117(50), 31902–31913. https://doi.org/10.1073/pnas.2018138117

[122] Stöckl, P., Hütter, E., Zwerschke, W., & Jansen-Dürr, P. (2006). Sustained inhibition of oxidative phosphorylation impairs cell proliferation and induces premature senescence in human fibroblasts. Experimental Gerontology, 41(7), 674–682. https://doi.org/10.1016/j.exger.2006.04.009

[123] Yoon, Y.-S., Byun, H.-O., Cho, H., Kim, B.-K., & Yoon, G. (2003). Complex II Defect via Down-regulation of Iron-Sulfur Subunit Induces Mitochondrial Dysfunction and Cell Cycle Delay in Iron Chelation-induced Senescence-associated Growth Arrest. Journal of Biological Chemistry, 278(51), 51577–51586. https://doi.org/10.1074/jbc.m308489200

[124] Chapman, J., Fielder, E., & Passos, J. F. (2019). Mitochondrial dysfunction and cell senescence: deciphering a complex relationship. FEBS Letters, 593(13), 1566–1579. https://doi.org/10.1002/1873-3468.13498

[125] Wiley, C. D., Velarde, M. C., Lecot, P., Liu, S., Sarnoski, E. A., Freund, A., … Campisi, J. (2016). Mitochondrial Dysfunction Induces Senescence with a Distinct Secretory Phenotype. Cell Metabolism, 23(2), 303–314. https://doi.org/10.1016/j.cmet.2015.11.011

[126] Bektas, A., Schurman, S. H., Sen, R., & Ferrucci, L. (2017). Human T cell immunosenescence and inflammation in aging. Journal of Leukocyte Biology, 102(4), 977–988. https://doi.org/10.1189/jlb.3ri0716-335r

[127] Sansoni, P., Vescovini, R., Fagnoni, F., Biasini, C., Zanni, F., Zanlari, L., … Passeri, M. (2008). The immune system in extreme longevity. Experimental Gerontology, 43(2), 61–65. https://doi.org/10.1016/j.exger.2007.06.008

[128] Kale, A., Sharma, A., Stolzing, A., Desprez, P.-Y., & Campisi, J. (2020). Role of immune cells in the removal of deleterious senescent cells. Immunity & Ageing, 17(1). https://doi.org/10.1186/s12979-020-00187-9

[129] Pereira, B. I., Devine, O. P., Vukmanovic-Stejic, M., Chambers, E. S., Subramanian, P., Patel, N., … Akbar, A. N. (2019). Senescent cells evade immune clearance via HLA-E-mediated NK and CD8+ T cell inhibition. Nature Communications, 10(1). https://doi.org/10.1038/s41467-019-10335-5

[130] Cai, Y., Zhou, H., Zhu, Y., Sun, Q., Ji, Y., Xue, A., … Deng, H. (2020). Elimination of senescent cells by β-galactosidase-targeted prodrug attenuates inflammation and restores physical function in aged mice. Cell Research, 30(7), 574–589. https://doi.org/10.1038/s41422-020-0314-9

Meet the Author:

Stellen Li is a junior majoring in Biostatistics and Mathematics with a minor in Statistics and Analytics. He took MATH 233 with Dr. Linda Green and found his interest in the field of topology. He is also involved as a research assistant at the Institute of Water at UNC, in which he conducts systematic literature reviews. He has also participated in Carolina Data Challenge (CDC) as well as mathematical modeling competitions to apply his knowledge in data analytics and statistical modeling. He hopes to conduct more research related to Biostatistics to further apply his interests in scientific research and statistics.

Polyhedra with hexagonal and triangular faces and three faces around each vertex

Linda Green [1*] and Stellen Li[2]

[1,2]University of North Carolina at Chapel Hill, Chapel Hill, NC, 27514, USA.

*Corresponding author(s). E-mail(s): greenl@email.unc.edu;
Contributing authors: stellen@ad.unc.edu;

Abstract

We analyze polyhedra composed of hexagons and triangles with three faces around each vertex, and their 3-regular planar graphs of edges and vertices, which we call "trihexes". Trihexes are analogous to fullerenes, which are 3-regular planar graphs whose faces are all hexagons and pentagons. Every trihex can be represented as the quotient of a hexagonal tiling of the plane under a group of isometries generated by 180° rotations. Every trihex can also be described with either one or three "signatures": triples of numbers that describe the arrangement of the rotocenters of these rotations. Simple arithmetic rules relate the three signatures that describe the same trihex. We obtain a bijection between trihexes and equivalence classes of signatures as defined by these rules. Labeling trihexes with signatures allows us to put bounds on the number of trihexes for a given number vertices $\boldsymbol{v}$ in terms of the prime factorization of $\boldsymbol{v}$ and to prove a conjecture concerning trihexes that have no "belts" of hexagons.

Keywords: Polyhedron, Fullerene, Hexagonal tiling, Three-regular graph, Two-faced map

1 Introduction

Motivated by the study of polyhedra, this paper analyzes 3-regular planar graphs whose faces all have three or six sides. We call these graphs *trihexes*. Trihexes have been analysed by Deza and Dutour ([1] and [2]), Grünbaum and Motzkin [3], and others. We refer to faces with three sides as "triangles" and faces with six sides as "hexagons", even though these faces may not have straight edges and may be unbounded.

Trihexes are analogous to fullerenes, which are 3-regular planar graphs whose faces all have five or six sides. Fullerenes have received much attention because when viewed as polyhedra, they have physical manifestations as carbon molecules. Fullerenes have been analyzed by Brinkmann, Goedgebeur, and McKay [4] and others.

In this paper, Section 3 explains how every triple of numbers (s, b, f) with $s \geq 0$, $b \geq 0$, and $0 \leq f \leq s$ (a "signature") describes a unique trihex. The number s gives the number of hexagons that lie in a chain capped by triangles (a "spine"), b gives the number of rings of hexagons ("belts") that surround and separate the spines, and f describes the rotation of the two spines relative to each other. Furthermore, every trihex can be described with at least one signature. Every trihex can also be described as the quotient of a hexagonal tiling of the plane under a group generated by 180° rotations, as shown in Section 4. In this context, the signatures (s, b, f) describe the arrangement of the rotocenters of these rotations. Although there can be three distinct signatures that describe the same trihex, simple arithmetic rules given in Section 5 relate the signatures that characterize the same trihex. We thus obtain a bijection between trihexes and equivalence classes of signatures as defined by these rules. In Section 7, we use our classification of trihexes in terms of signatures to put bounds on the number of trihexes with v vertices in terms of the prime factorization of $\frac{v}{4}$. In Section 8 we prove a conjecture about the "graph of curvatures" from [1].

The results in this paper can be applied to polyhedra whose faces are all triangles and hexagons and have three faces around each vertex, but are not necessarily convex. Trihexes with signatures $(0, b, 0)$ for $b \geq 1$ can be realized as the skeletons of non-convex polyhedra. All other trihexes can be realized as the skeletons of convex polyhedra. The correspondence between trihexes and convex polyhedra follows from Steinitz's theorem [5] or [6], as explained in Section 6.

2 Definitions and Preliminaries

Definition 1. *A* trihex *is a finite, connected, 3-regular planar graph whose faces all have three or six sides.*

Definition 2. *A* polyhedron *is a union of polygons in* $\mathbb{R}^3$ *which is homeomorphic to a sphere. Any pair of polygons intersect either in the empty set, a vertex, an edge, or a union of vertices and/or edges.*

Definition 3. *Two polyhedra are* equivalent *if there is an orientation preserving homeomorphism of the sphere that takes the faces, edges, and vertices of one polyhedron to the faces, edges, and vertices, respectively, of the other.*

The requirement that the homeomorphism be orientation preserving means that left-handed and right-handed versions of chiral polyhedra are not equivalent. We make the same distinction for trihexes. By a theorem of Whitney ([7], or see [8]) two planar graphs are isomorphic if and only if there is a homeomorphism of the sphere whose restriction to the planar graph gives a graph isomorphism. We consider trihexes equivalent if and only if an orientation-preserving homeomorphism can be found.

Definition 4. *Two trihexes are* equivalent *if they are not only isomorphic as graphs but if there is also an orientation-preserving homeomorphism of the plane that takes one graph to the other.*

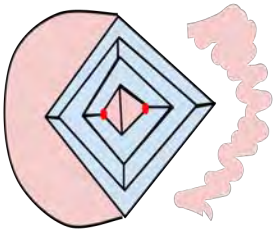


Fig. 1: Godseye with three pairs of hexagons.

Deza and Dutour ([1] and [2]) describe a family of 2-connected trihexes denoted by G_n or T_n, where n is half the number of hexagons. We will refer to these trihexes as godseyes, after the woven yarn craft figure that they resemble.

Definition 5. *A* godseye *is a trihex that consists of two adjacent triangles, surrounded by one or more nested pairs of hexagons, with two more adjacent triangles on the outside. The hexagons in each nested pair meet along opposide sides. See Figure 1.*

A standard Euler characteristic argument shows that every trihex has exactly four triangular faces. See, for example, [3]. The argument is as follows. Let f_6 be the number of hexagons in the trihex and f_3 be the number of triangular faces. The number of faces is $F = f_6 + f_3$. The number of edges is $E = \frac{6f_6 + 3f_3}{2}$, since each hexagonal face has six edges, each triangular face has three edges, and each edge is shared by two faces. The number of vertices is $V = \frac{6f_6 + 3f_3}{3}$, since each hexagonal face has six vertices, each triangular face has three vertices, and each vertex is shared by three faces. By Euler's formula, we have $V - E + F = 2$. Therefore, $\frac{6f_6 + 3f_3}{3} - \frac{6f_6 + 3f_3}{2} + f_6 + f_3 = 2$, which implies $f_3 = 4$.

Euler's formula places no restrictions on the number of hexagonal faces; however, Grünbaum and Motzkin showed that only even numbers of hexagonal faces can be achieved [3].

3 Building Trihexes From Spines and Belts

In this section, we describe ways to construct trihexes out of strings of hexagons capped by triangles ("spines"), possibly with rings of hexagons ("belts") separating the spines. Our construction echoes the construction given by Grünbaum and Motzkin in [3] but adds the consideration of "offset" defined below.

Definition 6. *A* belt *is a circuit of distinct hexagonal faces in a trihex such that each hexagon is adjacent to its neighbors on opposite edges [1].*

Definition 7. *A* spine *is a collection of distinct faces in a trihex $F_0, F_1, \cdots, F_{s+1}$, with $s \geq 0$, such that*

1. *F_0 and F_{s+1} are triangles,*
2. *$F_1, F_2, \cdots, F_s$ are hexagons, and*

Fig. 2: Spine of length 4.

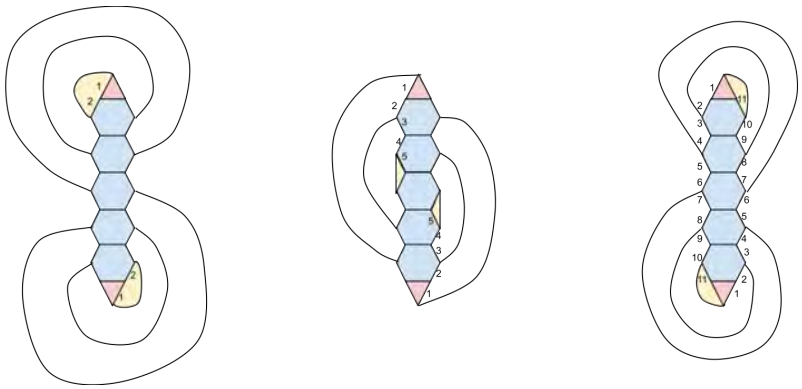
3. *For each hexagon F_i, $1 \leq i \leq s$, F_i is adjacent to F_{i-1} and to F_{i+1} along opposite edges of F_i.*

The internal edges *of the spine are the edges shared by F_i and F_{i+1} for $0 \leq i \leq s$ and the* external edges *of the spine are all the other edges. The* length *of the spine is the number s of hexagonal faces between the triangular faces. Note that a spine of length 0 is a pair of triangles that share an edge. We refer to the triangle F_0 as the* head triangle *of the spine and the triangle F_{s+1} as the* tail triangle*. Note that which triangle is considered the head triangle and which is considered the tail triangle depends only on the choice of numbering. The* head vertex *of the spine is the "tip" vertex of the head triangle, that is, the vertex that is not on an internal edge. The* tail vertex *of the spine is the "tip" vertex of the tail triangle.*

A trihex can be created from two spines of length s by attaching them along the $4s + 4$ external edges in each of their boundaries. This can be done in multiple ways. See Figure 3 for examples with $s = 5$.

Definition 8. *Suppose that two spines of length s are identified along their $4s + 4$ external edges. Starting with the head or tail vertex of one spine, travel counterclockwise around the boundary edges of this spine, until either a head or a tail vertex of the other spine is encountered, and count the number of edges traversed. We say that the two spines are attached with* offset $i \mod (s + 1)$ *if the number of edges traversed is $2i + 1$.*

To see that offset is well-defined, first note that the head vertex (or tail vertex) of the second spine must be identified to a vertex of the first spine where two faces of the first spine already meet. Otherwise, the trihex would not be 3-regular. Therefore, the number of edges traversed between the head or tail vertex of the first spine and a head or tail vertex of the second spine must be an odd number and is represented in the form $2i + 1$ for some integer i. In addition, since each spine has head and tail vertices that are $2s + 2$ edges apart, the number of edges traversed, going counterclockwise, to get from the *head* vertex of the first spine to any head or tail vertex of the second spine



(a) Offset 0 (b) Offset 2 (c) Offset 5

Fig. 3: Attaching spines

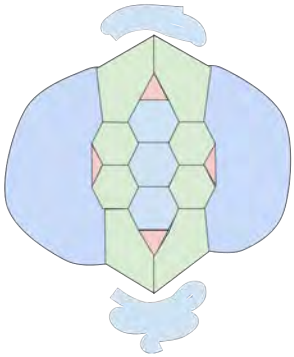
will be the same number $\mod (2s+2)$. It will also be the same number $\mod (2s+2)$ as the number of edges traversed to get from the *tail* vertex of the first spine to any head or tail vertex of the second spine. Since $2i+1 \equiv 2j+1 \mod (2s+2)$, if and only if $i \equiv j \mod (s+1)$, the offset is well-defined $\mod (s+1)$, no matter which head and tail vertices are used. It makes no difference which spine is considered the first spine and which is considered the second, since the same paths of edges are traversed whether traveling counterclockwise around one spine or the other, when going between a head or tail vertex of one spine and a head or tail vertex of the other.

For integers $s \geq 0$ and $b \geq 1$, we can also build a trihex out of a pair of spines of length s together with b belts of $2s+2$ hexagons, where the belts lie in between the two spines and encircle each spine. See Figure 4. As before, there are a variety of ways to attach the second spine onto the outermost belt, depending on where the head triangle of the second spine is inserted. Again, these different insertion points can be characterized by offsets.

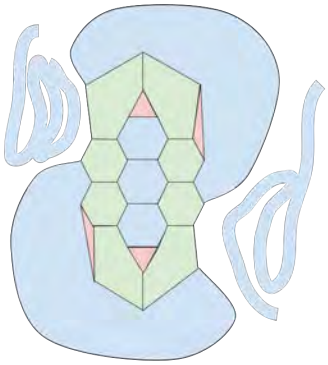
Suppose first that we have only one belt. Suppose we delete the belt and slide the two spines towards each other to fill in the space. If we slide them straight towards each other, along the edges of the belt that previously separated them, and then shift each spine slightly, either clockwise or counterclockwise around the other spine, then we form a new trihex with no belts between the spines. See Figure 5. Using the convention that we always shift clockwise, we can define the offset of the original trihex to be the offset of the new trihex with no belts between the spines (after shifting clockwise).

To find the offset when there are additional belts between the spines, we repeat the process of deleting belts and shifting the remaining pieces, starting from the outermost belt and working in.

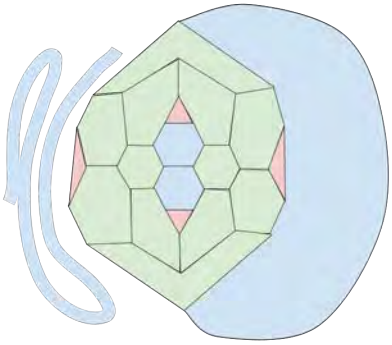
Definition 9. *When there are one or more belts between the spines, successively delete the belts, starting from the outermost belt, each time shifting one spine clockwise around the other spine. The offset of the original trihex is defined to be the offset of the resulting trihex that has no belts between the spines.*



(a) Two spines of length 3 with 1 belt and offset 1.



(b) Two spines of length 3 with 1 belt and offset 2.



(c) Two spines of length 2 with 2 belts and offset 0.

Fig. 4: Spines with belts between them.

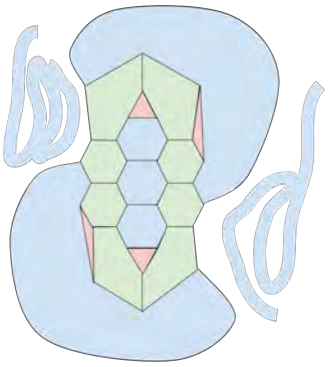
Note that shifting one spine clockwise around the other spine gives the same configuration as shifting the other spine clockwise around the first spine. Therefore, offset is well-defined irrespective of which spine is shifted with respect to the other and which belt is considered outermost vs. innermost.

For example, the original trihex in Figure 6 has offset 0.

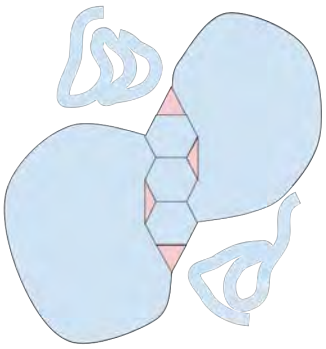
Definition 10. *Let $s \geq 0$, $b \geq 0$, and $0 \leq f \leq s$. If a trihex can be formed from two spines of length s with b belts between them and with offset f, then the* signature *of the trihex is the ordered triple (s, b, f).*

It is clear from the contruction that any two trihexes built from two spines of length s, with b belts between them and offset f are equivalent. In addition, Grünbaum and Motzkin [3] show that any trihex can be decomposed into spines and surrounding belts. Therefore, any trihex can be described with a signature (s, b, f) for some $s \geq 0$, $b \geq 0$, and $0 \leq f \leq s$.

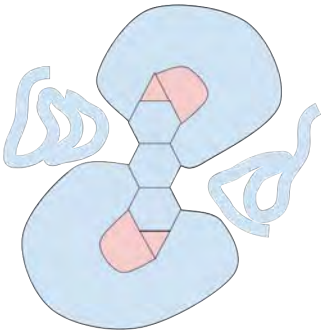
We summarize these facts in the following:



(a) The original trihex with one belt.



(b) After deleting the belt and shifting clockwise, the convention that we use. The offset of this trihex is 2, and therefore the offset of the original trihex is also 2.

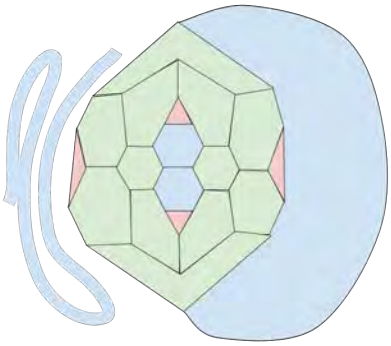


(c) After deleting the belt and shifting counterclockwise.

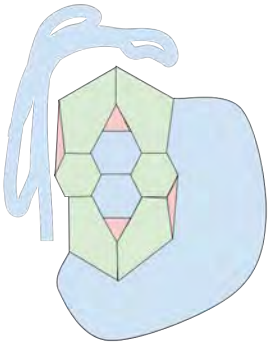
Fig. 5: Shifting clockwise vs. counterclockwise.

Theorem 1. *1. Given $s \geq 0$, $b \geq 0$, $0 \leq f \leq s$, there exists a trihex with signature (s, b, f).*
2. Any two trihexes with the same signature are equivalent.
3. Any trihex can be described with a signature (s, b, f) for some $s \geq 0$, $b \geq 0$, and $0 \leq f \leq s$.

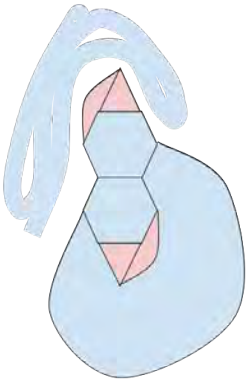
The signature for a trihex is not unique: decomposing a trihex in different ways can produce three different signatures, as detailed in Section 5.



(a) The original trihex with two belts.



(b) After deleting the one belt and shifting clockwise.



(c) After deleting the the second belt and shifting clockwise again. The offset of this trihex is 0, and so the offset of the original trihex is 0.

Fig. 6: Shifting to find offset.

4 Trihexes and Hexagonal Grid Coverings

In this section, we create a hexagonal tiling of the plane that covers a given trihex. Doing so allows us to develop rules for finding alternative signatures for a trihex in Section 5.

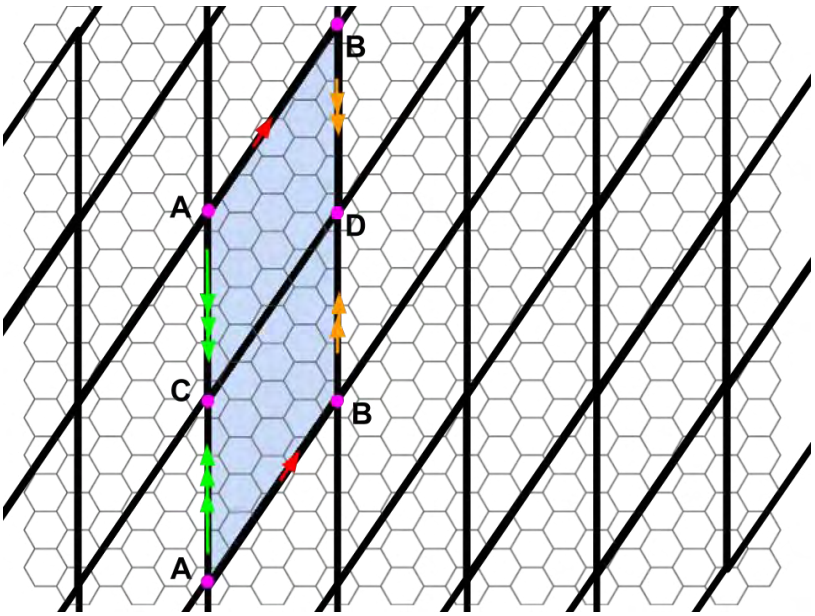


Fig. 7: Hexagonal grid with fundamental domain.

Consider a hexagonal tiling of the plane, made up of regular hexagons arranged in vertical strips such that two sides of each hexagon are horizontal. Superimpose a grid of parallelograms on the hexagonal tiling, such that each vertex of each parallelogram lies in the center of a hexagon, and each parallelogram has two vertical sides. See Figure 7.

Consider the group of isometries of the plane generated by 180° rotations around each vertex of the parallelogram grid. Form a quotient space (an orbifold) by identifying all points in the same orbit of this isometry group. A fundamental domain for this isometry group can be given by a pair of adjacent parallelograms, like the two parallelograms shaded blue in Figure 7, as explained below.

The orbits of points in this pair of parallelograms cover the entire plane, since a rotation around point C and then around the upper point marked A translates the pair of parallelograms up, and a rotation around C and then around the lower point marked A translates the parallelograms down. Repeating these pairs of rotations translates the parallelograms over an entire vertical strip. A rotation around C moves and inverts this vertical strip to cover the strip to the left, while a rotation around D covers the strip to the right. Repeated alternating rotations around points C and D covers all additional vertical strips to the left and the right.

No smaller subset of this double parallelogram region has orbits that cover the entire plane, by the following reasoning. A product of two 180° rotations is a translation through a vector twice the length of the vector connecting the rotocenters. A product of translations through two vectors is a translation through the sum of the vectors. Therefore, a product of an even number of 180° rotations around parallelogram grid vertices is a translation through a vector that is some sum $2m\vec{AB} + 2n\vec{AC}$

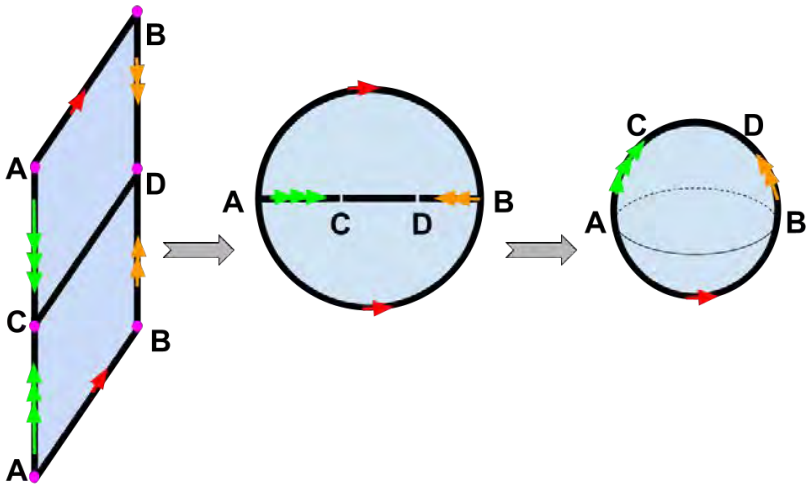


Fig. 8: Identifying edges creates a topological sphere.

for some integers m and n. A product of an 180° rotation and a translation is a 180° rotation whose rotocenter is the original rotocenter shifted by half the translation vector. Therefore, a product of an odd number of 180° rotations around parallelogram vertices is a 180° rotation around a rotocenter that is a parallelogram grid vertex shifted by $\frac{1}{2}(2m\vec{AB} + 2n\vec{AC})$ for some integers m and n, which is just another parallelogram grid vertex. The points interior to the double parallelogram region cannot be transformed onto each other by 180° rotations around parallelogram grid vertices or translations by linear combinations of $2\vec{AB}$ and $2\vec{AC}$. Therefore, the double parallelogram region is a minimal size region whose orbits cover the plane, i.e. a fundamental domain.

Although no points in the interior of the fundamental domain are identified under the isometry group, many pairs of points on the edges of the fundamental domain are identified with each other. This is indicated by the arrows in Figure 7: a rotation around point D identifies the edges above and below D, a rotation around C identifies the edges above and below C, and rotation around A followed by rotation around C identifies the top and bottom edges between the points marked A and B. After identifying edges, the resulting quotient is a topological sphere. See Figure 8.

Note that the hexagonal tiling is preserved by all of the 180° rotations. Therefore, it can be projected via the quotient map down to the quotient sphere. Each hexagon that lies entirely inside the fundamental domain will project onto a hexagon on the sphere. The partial hexagons that are cut off by edges of the fundamental domain, but do not contain the vertices marked A, B, C, and D, attach up in pairs and therefore also project onto hexagons in the quotient sphere. The partial hexagons that contain

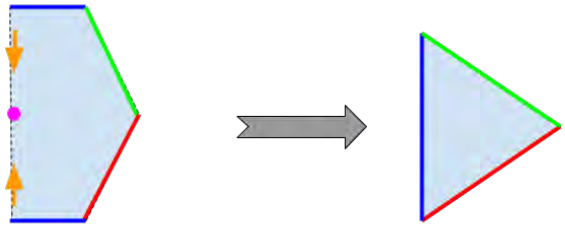


Fig. 9: The quotient of a hexagon under a 180° rotation around its center point.

the vertices marked A, B, C, and D get identified in such a way that only half a hexagon appears in the quotient sphere. This half hexagon is has its diameter identified by a 180° rotation, so that it forms a triangle in the quotient sphere. See Figure 9. Therefore, the hexagonal tiling naturally forms a pattern of hexagons and triangles on the quotient sphere, with three faces around each vertex. So all the requirements for a trihex are satisfied.

A signature for a trihex described as the quotient of a hexagonal tiling can be read off directly from the tiling, as illustrated in Figures 10 and 11. Each vertical string of hexagons between vertices labeled A and C, together with the hexagons centered at A and C, projects to one spine, and each vertical string of hexagons between vertices labeled B and D, together with the hexagons centered at B and D, projects to a second spine. The hexagons that lie in vertical strips between the vertical sides of a fundamental domain project to belts between the two spines. Therefore the number s in the signature (s, b, f) is the number of hexagons that lie in a vertical strip strictly between the hexagons at vertex A and vertex B. The number b is the number of vertical columns of hexagons in the tiling that lie entirely between the two vertical edges of the fundamental domain.

To find the offset for the trihex, choose a hexagon centered at a vertex A and translate it along the diagonal strip of hexagons in the approximately southwest (SW) to northeast (NE) direction, until it coincides with a hexagon in the vertical column of hexagons containing vertices B and D. If we hit a hexagon that is k hexagons below a vertex B or D, then our trihex will have offset k. This is because a hexagon at vertex A projects to a head triangle in the trihex, and translating this hexagon one column to the right in the SW to NE direction corresponds to deleting one belt and shifting the head triangle clockwise in the trihex. The ultimate position of the hexagon at vertex A after translating all the way to the right edge of the fundamental domain corresponds to the ultimate position that the head vertex of the head triangle is inserted along the second spine in the trihex, which is the offset. For example, the fundamental domain shown in Figure 11 covers the trihex with signature $(4, 3, 3)$.

It is now possible to conclude the following:

Theorem 2. *Every trihex can be produced as the quotient of a hexagonal tiling of the plane under a group of isometries generated by* 180° *rotations around the vertices of a superimposed parallelogram grid.*

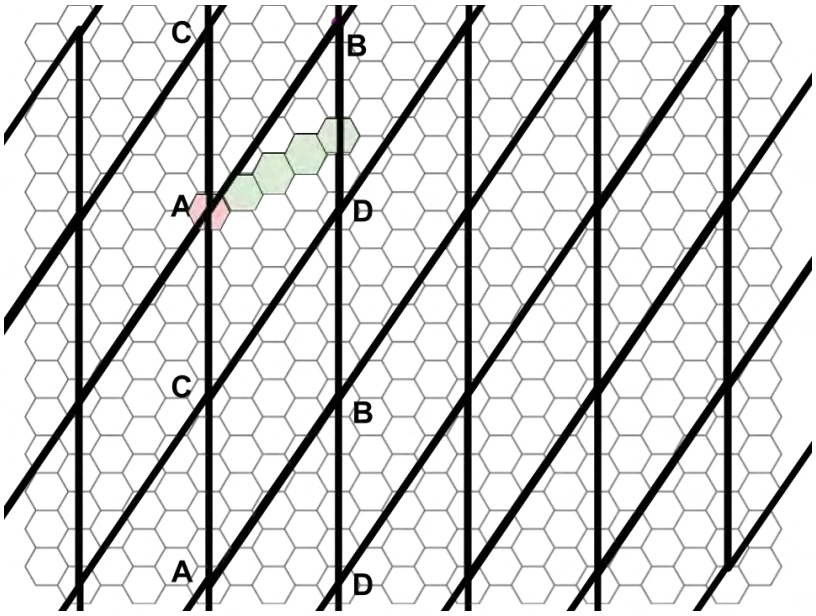


Fig. 10: Calculating offset.

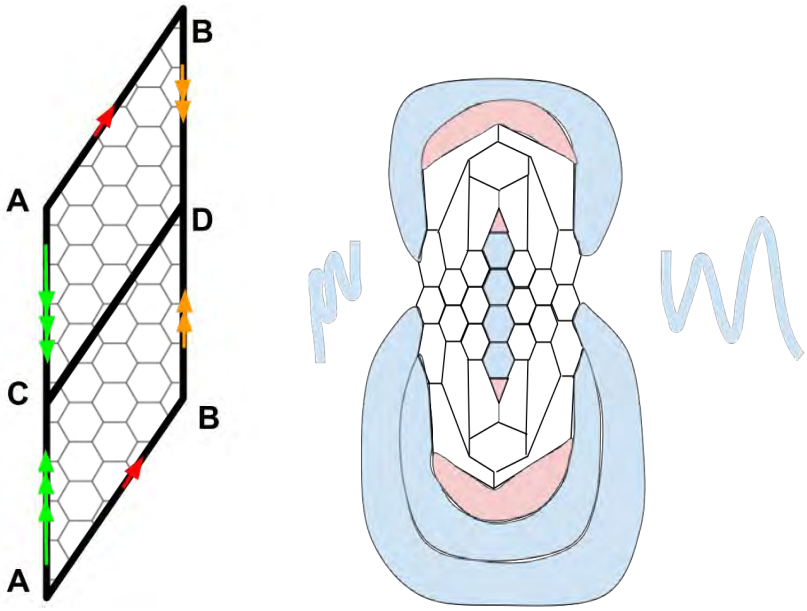


Fig. 11: The fundamental domain that covers the trihex with signature $(4, 3, 3)$.

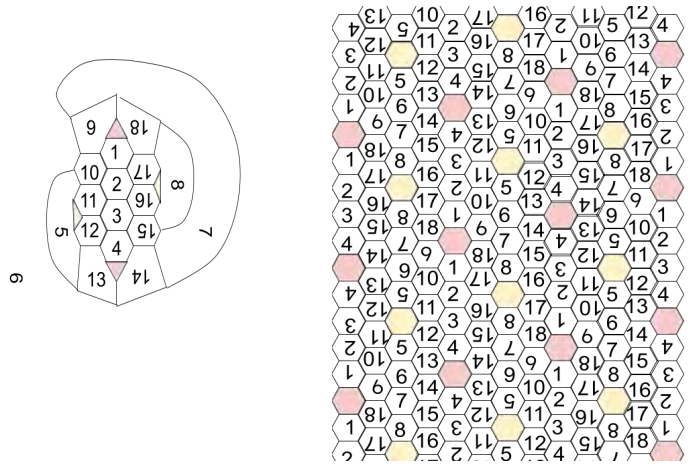


Fig. 12: The trihex with signature $(4, 1, 2)$ and the hexagonal tililng that covers it.

Proof. Recall from Theorem 1 that any trihex can be described with a signature (s, b, f) with $s \geq 0$, $b \geq 0$, and $0 \leq f \leq s$. For any such triple of integers (s, b, f), build a hexagonal tiling with a superimposed parallelogram grid as follows. Start with a tiling of the plane by regular hexagons in which two sides of each hexagon are horizontal. Put two vertices of a parallelogram on the centers of two hexagons in the same vertical column that are separated by s hexagons strictly between them. Put the other two vertices of the parallologram on the centers of hexagons in another vertical column, $b + 1$ columns to the right of the first column. This second pair of vertices should also be separated by s hexagons strictly between them. Shift the second pair of vertices up or down as needed, so that when the hexagons containing the first pair of vertices are translated along a SW to NE diagonal, through $b+1$ columns of hexagons, they end up f hexagons below the hexagons occupied by the second pair of vertices. We now have one parallelogram whose vertices lie on the centers of hexagons. Tile the plane with translated copies of this parallelogram to create a parallelogram grid. The quotient of the hexagonal tiling by the group generated by 180° rotations around parallelogram vertices is a trihex with signature (s, b, f).

□

The process of creating a hexagonal tiling that covers a given trihex can be thought of as "unwrapping" the trihex around each triangle. Figure 12 shows the unwrapping of the trihex $(4, 1, 2)$. Numbered hexagons in the trihex correspond to numbered hexagons in the hexagonal tiling.

5 Equivalent Signatures

The signature (s, b, f) for a trihex is not necessarily unique. This section develops rules for finding alternative signatures for a trihex based on a given signature and shows

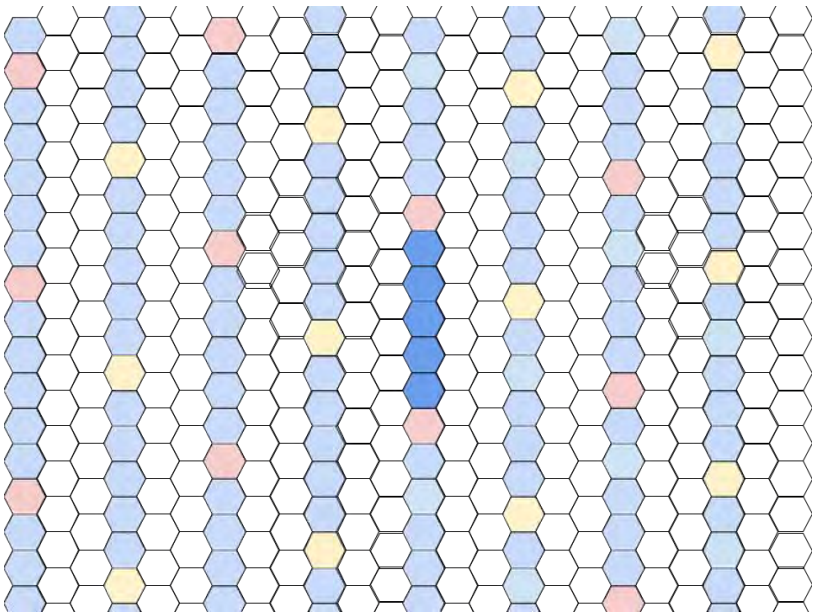


Fig. 13: A hexagonal tiling for the trihex $(5, 2, 2)$.

that there is a one to one correspondence between equivalence classes of signatures, as defined in this section, and equivalence classes of trihexes, as defined in Section 2.

Start with a trihex with signature (s_1, b_1, f_1). For example, a hexagonal tiling for trihex $(5, 2, 2)$ is shown in Figure 13. Vertical spines are shaded blue. Special hexagons are shaded pink and yellow. The superimposed parallelogram grid is not drawn.

There are two additional ways to describe the trihex. Instead of using vertical strings of hexagons to make spines, we could build spines by setting off from a special hexagon at an angle 60° clockwise from due north or at an angle 120° clockwise from due north. We will refer to these directions as the southwest (SW) to northeast (NE) direction and the northwest (NW) to southeast (SE) direction. See Figures 14 and 15. We will use the notation (s_2, b_2, f_2) to refer to the signature when we build spines in the SW to NE direction and (s_3, b_3, f_3) to refer to the signature when we go in the NW to SE direction.

Suppose we go from SW to NE. See Figure 14, where four special hexagons are labelled T, B, L, R (top, bottom, left, and right). Since the original offset was f_1, this means that if we start at a special hexagon, say L, and translate it through $b_1 + 1$ vertical columns of hexagons, always along a SW to NE diagonal of hexagons, thereby arriving in another vertical column with special hexagons, we land f_1 hexagons below a special hexagon. For each additional $b_1 + 1$ vertical columns of hexagons we go through in the SW to NE direction, we land an additional f_1 hexagons below a special hexagon. If at any moment we land a multiple of $s_1 + 1$ hexagons below a special hexagon, then we are directly on a special hexagon, since special hexagons appear every $s_1 + 1$ hexagons in the vertical column. Let j_2 be the smallest integer ≥ 1 such that $j_2 \cdot f_1$ is a multiple of $s_1 + 1$ (i.e. j_2 is the order of f_1 in Z_{s_1+1}). Then the first time that we land

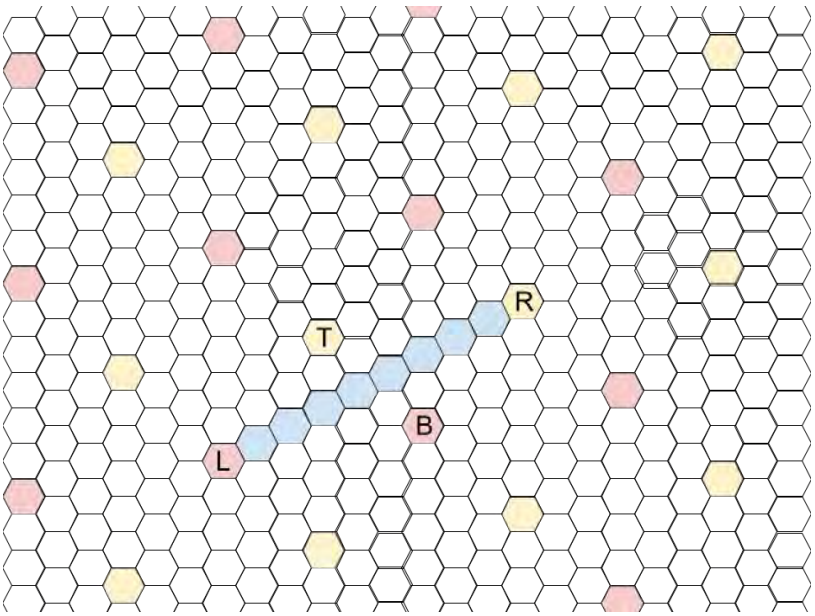


Fig. 14: An alternative spine of (5,2,2) with length 8.

directly on a special hexagon is when we have traveled through $j_2 \cdot (b_1 + 1)$ vertical columns. The string of hexagons in the SW to NE diagonal that connects the original special hexagon to this final special hexagon projects to a spine in the quotient trihex. This spine will contain $j_2(b_1 + 1) - 1$ hexagons, since the final hexagon projects to a triangle. So this spine has length $s_2 = j_2(b_1 + 1) - 1$. For example, if we start at a special hexagon of a $(5, 2, 2)$ hexagonal grid and head northeast, we will create spines of length 8, because $b_1 + 1 = 3$, $s_1 + 1 = 6$, $f_1 = 2$, the order of 2 in Z_6 is $j_2 = 3$, and $j_2 \cdot (b_1 + 1) - 1 = 3 \cdot 3 - 1 = 8$. See Figure 14.

Suppose instead that we translate a special hexagon in the NW to SE direction. See Figure 15. If we travel through $b_1 + 1$ vertical columns of hexagons, thereby arriving in another vertical column with special hexagons, we land $f_1 + b_1 + 1$ hexagons below a special hexagon, since each translation through a single column in the NW to SE direction puts us one hexagon below where we would move to when translating in the SW to NE direction. For each additional $b_1 + 1$ vertical columns of hexagons we go through in the NW to SE direction, we land an additional $f_1 + b_1 + 1$ hexagons below a special hexagon. So the first time we hit a special hexagon is when we have traveled through $j_3(b_1 + 1)$ hexagons, where j_3 is the smallest integer ≥ 1 such that $j_3(f_1 + b_1 + 1)$ is a multiple of $s_1 + 1$, i.e. j_3 is the order of $f_1 + b_1 + 1$ in Z_{s_1+1}. At this point we will have created a spine of length $j_3(b_1 + 1) - 1$. So $s_3 = j_3(b_1 + 1) - 1$. For example, if we start with $(s_1, b_1, f_1) = (5, 2, 2)$, then $f_1 + b_1 + 1 = 5$ and $s_1 + 1 = 6$, and 5 has order $j_3 = 6$ in Z_6. Since $j_3(b_1 + 1) - 1 = 6 \cdot 3 - 1 = 17$, we will have a spine of length 17. See Figure 15.

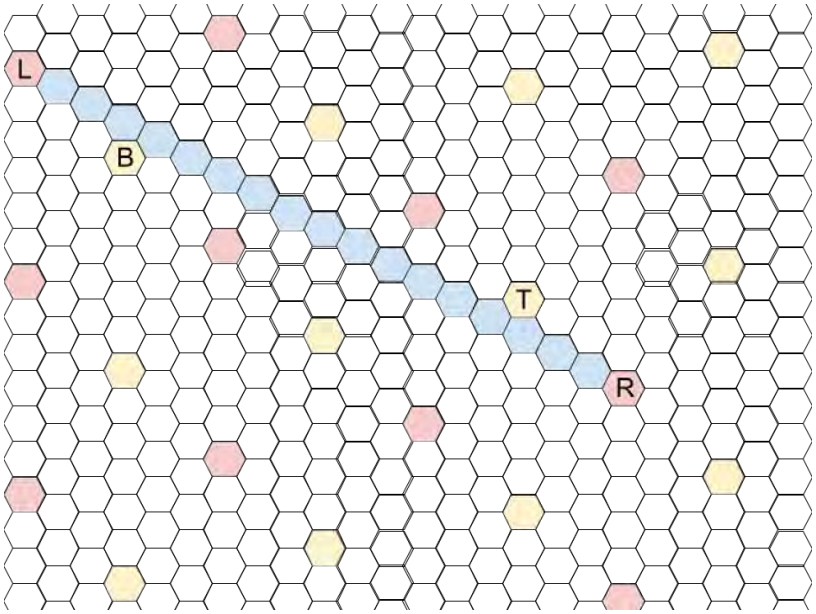


Fig. 15: An alternative spine for (5,2,2) with length 17.

To find the number of belts in the SW to NE decomposition, note that the total number h of hexagons, based on the original signature of (s_1, b_1, f_1), is given by $h = 2s_1 + b_1(2s_1 + 2) = 2s_1b_1 + 2s_1 + 2b_1$, since each of the two spines contains s_1 hexagons and each of the b_1 surrounding belts contains $2s_1 + 2$ hexagons. If (s_2, b_2, f_2) is the new signature based on spines in the SW to NE direction, then h must also equal $2s_2b_2 + 2s_2 + 2b_2$. So $b_2 = \dfrac{h - 2s_2}{2s_2 + 2}$. Similarly, the number of belts for the NW to SE decomposition with signature (s_3, b_3, f_3) is given by $b_3 = \dfrac{h - 2s_3}{2s_3 + 2}$. For example, for the trihex $(5, 2, 2)$, we have $h = 2 \cdot 5 \cdot 2 + 2 \cdot 5 + 2 \cdot 2 = 34$. We saw that $s_2 = 8$ and $s_3 = 17$. So the number of belts using the SW to NE spines is $b_2 = \dfrac{34 - 2 \cdot 8}{2 \cdot 8 + 2} = \dfrac{18}{18} = 1$. The number of belts using the NW to SE spines is $b_3 = \dfrac{34 - 2 \cdot 17}{2 \cdot 17 + 2} = \dfrac{0}{36} = 0$. Note that for the SW to NE decomposition, the belts in the trihex are covered by diagonal strips of hexagons in the SW to NE direction that lie between the diagonal strips containing special hexagons. Similarly, for the NW to SE decomposition, the belts in the trihex are covered by diagonal strips of hexagons in the NW to SE direction. We will call these diagonal strips of hexagons "belt strips".

To find the offset for the SW to NE signature, we first need to find a special hexagon that is adjacent to the belt strips around a SW to NE spine. Label the special hexagons on the left and right ends of a fixed diagonal spine L and R, respectively. Label the special hexagons that are adjacent to the belt strips T and B, where T is adjacent on top and B is adjacent on bottom. See Figure 14. Hexagons T and B

project to the head and tail triangles of a second SW to NE spine whose position relative to the first SW to NE spine will give the offset. Hexagon T will be a special hexagon that is $b_2 + 1$ hexagons above the original diagonal spine, or equivalently, the diagonal spine is $b_2 + 1$ hexagons below T. Recall that each time we travel $b_1 + 1$ columns along the diagonal spine in the SW to NE direction, we land an additional f_1 hexagons below a special hexagon. Therefore, we need to find a number p_2 such that $p_2 \cdot f_1 \equiv (b_2 + 1) \mod (s_1 + 1)$, and travel $p_2(b_1 + 1)$ columns to the right, in order to land in the same column as T, but $b_2 + 1$ hexagons below it. For simplicity, pick p_2 to the smallest number ≥ 1 such that $p_2 \cdot f_1 \equiv (b_2 + 1) \mod (s_1 + 1)$.

To find the corresponding offset, notice that since T is $p_2(b_1 + 1)$ columns to the right of L's column, it will be $(s_2 + 1) - p_2(b_1 + 1)$ columns to the left of R's column. If $b_2 = 0$, the number of columns to the left of R is one more than the offset, so $f_2 = (s_2 + 1) - p_2(b_1 + 1) - 1 = s_2 - p_2(b_1 + 1)$. If $b_2 > 0$, then calculating offset involves deleting b_2 diagonal belts in the quotient trihex and moving clockwise, which is equivalent to moving b_2 columns to the right in the **NW to SE** direction in the hexagonal tiling of the plane. Therefore, the offset will be b_2 smaller, that is, $f_2 = s_2 - p_2(b_1 + 1) - b_2 \mod (s_2 + 1)$. Note that offset is defined mod $(s_2 + 1)$ since s_2 is the length of the diagonal spine. See Figure 14.

For the $(5, 2, 2)$ trihex, $s_1 = 5$, $b_1 = 2$, $f_1 = 2$, $s_2 = 8$, and $b_2 = 1$. The number p_2 is defined as the smallest number ≥ 1 such that $p_2 \cdot f_1 \equiv (b_2 + 1) \mod (s_1 + 1)$, i.e. such that $p_2 \cdot 2 \equiv 2 \mod 6$. Therefore, $p_2 = 1$, and $f_2 = 8 - 1 \cdot 3 - 1 \mod 9 = 4$. The $(5, 2, 2)$ trihex has an alternative signature of $(8, 1, 4)$.

To find the offset for the NW to SE signature, label the special hexagons on the left and right ends of the diagonal spine in the NW to SE direction with L and R, respectively, and the special hexagons that are adjacent to the surrounding belt strips T and B, where T is adjacent on top and B is adjacent on bottom. See Figure 15. Hexagon T will be $b_3 + 1$ hexagons above the diagonal spine, or equivalently, the diagonal spine is $b_3 + 1$ hexagons below T. Recall that each time we travel $b_1 + 1$ columns along the diagonal spine in the NW to SE direction, we land an additional $f_1 + b_1 + 1$ hexagons below a special hexagon. Therefore, we need to find a number p_3 such that $p_3 \cdot (f_1 + b_1 + 1) \equiv (b_3 + 1) \mod (s_1 + 1)$, and travel $p_3(b_1 + 1)$ columns to the right, in order to land in the same column as T, but $b_3 + 1$ hexagons below it. For simplicity, pick p_3 to the smallest number ≥ 1 such that $p_3 \cdot (f_1 + b_1 + 1) \equiv (b_3 + 1) \mod (s_1 + 1)$.

To find the corresponding offset, notice that since T is $p_3(b_1 + 1)$ columns to the right of L's column, it will be $(s_3 + 1) - p_3(b_1 + 1)$ columns to the left of R's column. If $b_3 = 0$, the number of columns to the left of R is equal to the offset, instead of one more than the offset, like it was for the SW to NE spine. So $f_3 = (s_3 + 1) - p_3(b_1 + 1)$. If $b_3 > 0$, then calculating offset in the quotient trihex involves deleting b_3 diagonal belts and moving clockwise, which is equivalent to simply moving the hexagon T straight down in its column in the hexagonal grid covering. Therefore, the offset will be $f_3 = (s_3 + 1) - p_3(b_1 + 1) \mod (s_3 + 1)$. Again, the offset is defined mod $s_3 + 1$ since s_3 is the length of the diagonal spine. See Figure 15.

For the $(5, 2, 2)$ trihex, $s_1 = 5$, $b_1 = 2$, $f_1 = 2$, $s_3 = 17$, and $b_3 = 0$, so $f_1 + b_1 + 1 = 5$. The number p_3 is defined as the smallest number ≥ 1 such that $p_3 \cdot (f_1 + b_1 + 1) \equiv b_3 + 1$

mod (s_1+1), i.e. such that $p_3 \cdot 5 \equiv 1 \mod 6$. Therefore, $p_3 = 5$, and $f_3 = 18 - 5 \cdot 3 \mod 18 = 3$. The $(5,2,2)$ trihex has an alternative signature of $(17,0,3)$.

Definition 11. *Given a trihex with signature (s_1, b_1, f_1), the* equivalent signatures *for this trihex are the original signature (s_1, b_1, f_1) along with signatures (s_2, b_2, f_2) and (s_3, b_3, f_3) found using the following algorithms.*

Using the SW to NE spine :

1. Find the smallest number $j_2 \geq 1$ such that $j_2 \cdot f_1 \equiv 0 \mod (s_1+1)$.
2. $s_2 = j_2(b_1+1) - 1$.
3. Compute the total number of hexagons in the original trihex: $h = 2s_1 \cdot b_1 + 2s_1 + 2b_1$.
4. $b_2 = \dfrac{h - 2s_2}{2s_2 + 2}$.
5. Find the smallest number $p_2 \geq 1$ such that $p_2 \cdot f_1 \equiv (b_2+1) \mod (s_1+1)$.
6. $f_2 = s_2 - p_2(b_1+1) - b_2 \mod (s_2+1)$.

Using the NW to SE spine (only steps 1, 5, and 6 are different):

1. Find the smallest number $j_3 \geq 1$ such that $j_3(f_1 + b_1 + 1) \equiv 0 \mod (s_1+1)$.
2. $s_3 = j_3(b_1+1) - 1$.
3. Compute the total number of hexagons in the original trihex: $h = 2s_1 \cdot b_1 + 2s_1 + 2b_1$.
4. $b_3 = \dfrac{h - 2s_3}{2s_3 + 2}$.
5. Find the smallest number $p_3 \geq 1$ such that $p_3 \cdot (f_1 + b_1 + 1) \equiv (b_3+1) \mod (s_1+1)$.
6. $f_3 = s_3 + 1 - p_3(b_1+1) \mod (s_3+1)$.

The signatures (s_1, b_1, f_1), (s_2, b_2, f_2), and (s_3, b_3, f_3) give the three alternative descriptions of the same arrangement of special hexagons on a hexagonal tiling of the plane, found by rotating the "vertical" direction by 0 or 180 degree, 60 or 240 degrees, and 120 or 300 degrees clockwise, respectively. Therefore, this definition of equivalent signatures does in fact describe an equivalence relationship. Although the three signatures are usually distinct, it is possible for all three to be the same. See Table 1. It is not possible for two of the three signatures to be the same and the third signature different: if two signatures are the same, say $(s_1, b_1, f_1) = (s_2, b_2, f_2)$, then rotating the hexagonally tiled plane by 60 degrees will produce the same configuration of special hexagons. Therefore, rotating a second time by the 60 degrees will again produce the same configuration of special hexagons, so (s_3, b_3, f_3) will also be the same.

Recall that two trihexes are considered equivalent if they are not only isomorphic as graphs, but if there is also an orientation-preserving homeomorphism of the plane that takes one graph to the other. Chiral trihexes that are mirror images of each other are not considered equivalent. Figure 16 illustrates the following relationship:

Proposition 3. *A trihex with signature (s, b, f) has a mirror image trihex with signature $(s, b, s - f - b \mod (s+1))$.*

Proof. The mirror image of trihex (s, b, f) will still have the same spine lengths as the original (s) and the same number of belts in between them (b). See Figure 16

Suppose that $b = 0$. If the original trihex has offset f, then its mirror image will have offset $s - f$. Suppose $b > 0$. Since the offset of the original trihex is f, if we delete the b belts one at a time and shift the head vertex clockwise each time, then the head

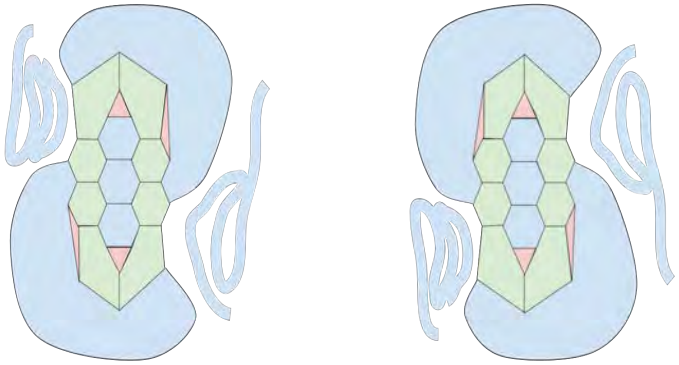


Fig. 16: The trihex $(3, 1, 2)$ and its mirror image $(3, 1, 0)$

vertex lands at offset f. Therefore, in the mirror image trihex, if we delete belts one at a time and shift the head vertex *counterclockwise* each time, then the head vertex will land at the mirror image position offset $s - f$. Since moving counterclockwise instead of clockwise increases offset by 1 for each belt that is removed, the actual offset for the mirror image found by shifting clockwise will be b less than $s - f$, i.e. $s - f - b \mod (s + 1)$. □

We can now state our classification of trihexes: trihexes are precisely indexed by the equivalence classes of triples (s, b, f) under the relations for triples stated in Definition 11.

Theorem 4. *1. Suppose T_0 and T_1 are trihexes with signatures (s_0, b_0, f_0) and (s_1, b_1, f_1). Then T_0 and T_1 are equivalent trihexes if and only if (s_0, b_0, f_0) and (s_1, b_1, f_1) are equivalent signatures.*

Therefore, there is a bijection between equivalence classes of trihexes and equivalence classes of signatures.

2. Suppose T_0 and T_1 are trihexes with signatures (s_0, b_0, f_0) and (s_1, b_1, f_1). Then T_0 and T_1 are isomorphic as graphs but not equivalent (i.e. they are mirror images of each other), if and only if (s_1, b_1, f_1) is equivalent to $(s_0, b_0, s_0 - f_0 - b_0 \mod (s_0 + 1))$.

Therefore there is a bijection between graph isomorphism classes of trihexes and sets of signatures that are either equivalent or mirror equivalent.

Proof. Suppose the two trihexes T_0 and T_1 are isomorphic as graphs. By a theorem of Whitney ([7], or see [8]), there is a homeomorphism of the sphere whose restriction to T_0 gives a graph isomorphism to T_1. Lift this homeomorphism to a map from the hexagonal tiling that covers T_0 to the hexagonal tiling that covers T_1. This lifted map is a homeomorphism of the hexagonally tiled plane that takes hexagons to hexagons and special hexagons to special hexagons. There is a unique isometry of the plane that agrees with the homeomorphism on all the vertices of the hexagonal tiling. The isometry is orientation preserving if and only if the original homeomorphism is.

If the isometry is orientation preserving, then it must be either a rotation by a multiple of 60°, or a translation, since these are the only orientation preserving

isometries of the plane that preserve the hexagonal grid. A rotation by 180° or 0° or a translation takes vertical spines to vertical spines. A rotation by 60° or 240° counterclockwise takes vertical spines to NW to SE spines, and a rotation by 120° or 300° counterclockwise takes vertical spines to SW to NE spines. Therefore, (s_0, f_0, b_0) must be equivalent to (s_1, f_1, b_1).

If the isometry is orientation reversing, then reflect the first hexagonal tiling through a vertical line centered at special hexagons. Consider the isometry from this reflected hexagonal tiling to the second hexagonal tiling formed as the composition of the reflection followed by the original isometry. This composition gives an orientation preserving isometry from the mirror image of the first hexagonal tiling to the second hexagonal tiling. Therefore, the reflected hexagonal tiling, whose signature is $(s_0, b_0, s_0 - f_0 - b_0 \mod (s_0 + 1))$, has signature equivalent to (s_1, b_1, f_1)..

Conversely, suppose the signature (s_1, b_1, f_1) is equivalent to the signature (s_0, b_0, f_0). By Theorem 2, both trihexes T_0 and T_1 arise as quotients of hexagonal tilings under groups of isometries generated by 180° rotations. Since (s_0, b_0, f_0) and (s_1, b_1, f_1) are equivalent signatures, the rotocenters of these rotations are the same, and so these isometry groups are the same. Therefore, the trihexes must be equivalent. If (s_1, b_1, f_1) is equivalent to $(s_0, b_0, s_0 - f_0 - b_0 \mod (s_0 + 1))$, then the grids of rotocenters for these rotations are mirror images of each other. Therefore, there is a reflection that takes one hexagonal tiling to the other, takes special hexagons to special hexagons, and is preserved by the action of the rotation groups. This reflection projects to an orientation-reversing homeomorphism between quotient spheres that is a graph isomorphism between T_0 and T_1.

The existence of bijections now follows from Theorem 1, which says that every signature is realized by a trihex and every trihex has a signature. □

Although each trihex has three equivalent signatures, in some cases, the signatures are repetitions of each other. For example, the alternative signatures for $(6, 0, 2)$ are $(6, 0, 2)$ and $(6, 0, 2)$. Whenever a trihex has a symmetry type that includes an order 3 rotation, then its three signatures will be repetitions of each other, because the order 3 rotation lifts to an order 3 or order 6 rotation of the hexagonal tiling that takes special hexagons in the vertical direction to special hexagons in the NW to SE direction or the SW to NE direction.

Table 1 gives the signatures for all trihexes with 20 hexagons or fewer (44 vertices or fewer). Each row gives the three equivalent signatures for a trihex. The three signatures are ordered so that the signature with the smallest value of b is on the left, with preference given to the signature with smaller value of f in case of a tie.

In this table, the rows with signatures $(b, 0, 0)$, $(b, 0, b)$, $(0, b, 0)$ for $b \geq 1$ correspond to godseyes. A godseye constructed from from two spines of length 0 with b belts of hexagons between them clearly has signature $(0, b, 0)$. Its other two signatures can be found by lifting the godseye to its hexagonal tiling. Since the godseye has spines of length 0, its hexagonal tiling has vertical columns filled with special hexagons, separated by b columns of (non-special) hexagons. Spines in the NW to SE and SW to NE directions have length b, with no belts between them, and offsets of 0 and b.

(s_1, b_1, f_1)	(s_2, b_2, f_2)	(s_3, b_3, f_3)	hexagons	vertices
(0, 0, 0)	(0, 0, 0)	(0, 0, 0)	0	4
(1, 0, 0)	(1, 0, 1)	(0, 1, 0)	2	8
(2, 0, 0)	(2, 0, 2)	(0, 2, 0)	4	12
(2, 0, 1)	(2, 0, 1)	(2, 0, 1)	4	12
(3, 0, 0)	(3, 0, 3)	(0, 3, 0)	6	16
(3, 0, 1)	(3, 0, 2)	(1, 1, 1)	6	16
(1, 1, 0)	(1, 1, 0)	(1, 1, 0)	6	16
(4, 0, 0)	(4, 0, 4)	(0, 4, 0)	8	20
(4, 0, 1)	(4, 0, 2)	(4, 0, 3)	8	20
(5, 0, 0)	(5, 0, 5)	(0, 5, 0)	10	24
(5, 0, 1)	(5, 0, 4)	(2, 1, 2)	10	24
(5, 0, 2)	(2, 1, 0)	(1, 2, 1)	10	24
(5, 0, 3)	(2, 1, 1)	(1, 2, 0)	10	24
(6, 0, 0)	(6, 0, 6)	(0, 6, 0)	12	28
(6, 0, 1)	(6, 0, 3)	(6, 0, 5)	12	28
(6, 0, 2)	(6, 0, 2)	(6, 0, 2)	12	28
(6, 0, 4)	(6, 0, 4)	(6, 0, 4)	12	28
(7, 0, 0)	(7, 0, 7)	(0, 7, 0)	14	32
(7, 0, 1)	(7, 0, 6)	(3, 1, 3)	14	32
(7, 0, 2)	(7, 0, 5)	(3, 1, 1)	14	32
(7, 0, 3)	(7, 0, 4)	(1, 3, 1)	14	32
(3, 1, 0)	(3, 1, 2)	(1, 3, 0)	14	32
(8, 0, 0)	(8, 0, 8)	(0, 8, 0)	16	36
(8, 0, 1)	(8, 0, 4)	(8, 0, 7)	16	36
(8, 0, 2)	(8, 0, 3)	(2, 2, 2)	16	36
(8, 0, 5)	(8, 0, 6)	(2, 2, 1)	16	36
(2, 2, 0)	(2, 2, 0)	(2, 2, 0)	16	36
(9, 0, 0)	(9, 0, 9)	(0, 9, 0)	18	40
(9, 0, 1)	(9, 0, 8)	(4, 1, 4)	18	40
(9, 0, 2)	(9, 0, 3)	(4, 1, 2)	18	40
(9, 0, 4)	(4, 1, 0)	(1, 4, 1)	18	40
(9, 0, 5)	(4, 1, 3)	(1, 4, 0)	18	40
(9, 0, 6)	(9, 0, 7)	(4, 1, 1)	18	40
(10, 0, 0)	(10, 0, 10)	(0, 10, 0)	20	44
(10, 0, 1)	(10, 0, 5)	(10, 0, 9)	20	44
(10, 0, 2)	(10, 0, 4)	(10, 0, 7)	20	44
(10, 0, 3)	(10, 0, 6)	(10, 0, 8)	20	44

Table 1: The three equivalent signatures for trihexes with 20 or fewer hexagons / 44 or fewer vertices.

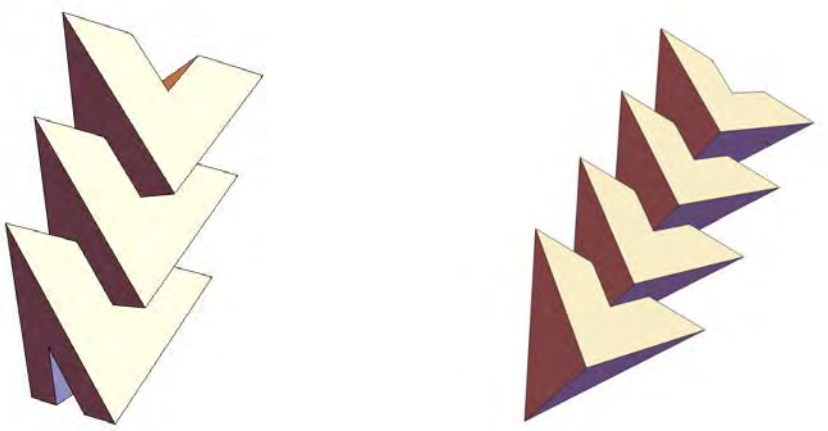
6 Convex vs. Non-convex Trihexes

The trihex signature determines whether it can arise from a convex polyhedron. We will first quote some preliminary facts.

Proposition 5. *1. Every trihex is 2-connected.*
2. If a trihex is not 3-connected, then it is a godseye.
3. A trihex is a simple graph.

Proof. Parts (i) and (ii) are proved in [1].

Part (iii): If a trihex had an edge loop, then it would have a face with only one edge. If it had a double edge, then it would either have a face with only two edges or



(a) Godseye with six pairs of hexagons (b) Godseye with seven pairs of hexagons

Fig. 17: Godseyes can be realized as non-convex polyhedra.

else each of the vertices on the double edge would be a separating vertex, contradicting Part (i). □

Theorem 6. *Any trihex with a signature* $(0, b, 0)$ *with* $b > 0$ *can be represented as the skeleton of a non-convex polyhedron, and it cannot be represented as the skeleton of a convex polyhedron. All other trihexes can be represented as skeletons of convex polyhedra.*

Proof. Steinitz's Theorem [5] or [6], says that a graph can be represented as the skeleton of a convex polyhedron if and only if the graph is simple, planar, and 3-connected. Trihexes with signature $(0, b, 0)$ with $b > 0$ are godseyes, which are not 3-connected: removing the two vertices shown in red in Figure 1 disconnects the graph. Therefore they are not the skeletons of convex polyhedra. However, they are the skeletons of non-convex polyhedra, as shown in Figure 17. All other trihexes are 3-connected, simple, planar graphs by Lemma 5. So by Steinitz's theorem, they are the skeletons of convex polyhedra. □

7 How Many Trihexes of Each Size?

Given $v \geq 0$, how many trihexes are there with v vertices? Let $\alpha(v)$ be the number of equivalence classes of trihexes with v vertices and let $\beta(v)$ be the number of graph isomorphism classes of trihexes with v vertices. These two quantities can be computed by counting signatures and accounting for duplicates, after establishing the following relationships, which are also evident in [3].

Lemma 1. *1. The triple* (s, b, f) *is a signature for a trihex with* h *hexagons if and only if* $s \geq 0$, $b \geq 0$, $0 \leq f \leq s$, *and* $\dfrac{h}{2} + 1 = (s + 1)(b + 1)$.

2. *The triple (s, b, f) is a signature for a trihex with v vertices if and only if $s \geq 0$, $b \geq 0$, $0 \leq f \leq s$, and $\frac{v}{4} = (s+1)(b+1)$.*
3. *Consequently, the number of hexagons in a trihex is even and the number of vertices of a trihex is divisible by 4.*

Proof. A trihex with signature (s, b, f) has $2s+b(2s+2) = 2(s+1)(b+1)-2$ hexagons, since each of the two spines contains s hexagons and each of the b belts contains $2s+2$ hexagons. So for $s \geq 0$, $b \geq 0$, $0 \leq f \leq s$, the triple (s, b, f) is a signature for a trihex with h hexagons if and only if $\frac{h}{2} = (s+1)(b+1) - 1$. Since each hexagon has six vertices and each of the four triangles in a trihex has four vertices, the number of vertices in a trihex is $v = \frac{6h+12}{3} = 2h + 4 = 4(s+1)(b+1)$. So the triple (s, b, f) with $s \geq 0$, $b \geq 0$, and $0 \leq f \leq s$ is a signature for a trihex with v vertices if and only if $\frac{v}{4} = (s+1)(b+1)$. Since $\frac{h}{2}$ and $\frac{v}{4}$ are integers, h is divisiby by 2 and v is divisible by 4. □

From this relationship, we can compute $\alpha(v)$ and $\beta(v)$ as follows. For each desired value of v, we can find the list of all triples (s, b, f) satisfying the inequalities $s \geq 0$, $b \geq 0$, and $0 \leq f \leq s$ and the equation $\frac{v}{4} = (s+1)(b+1)$. Then we can test which triples satisfy the relationships in Definition 11 and therefore represent equivalent trihexes. Taking these equivalences into account yields $\alpha(v)$. Checking for triples that satisfy the mirror equivalence relation in Theorem 4 yields $\beta(v)$. See Table 2 for counts of $\alpha(v)$ and $\beta(v)$ for $v \leq 200$.

Let $\sigma(v)$ be the number of triples that form a signature for a trihex with v vertices; that is, the number of triples (s, b, f) with $s \geq 0$, $b \geq 0$, $0 \leq f \leq s$, and $v = 4(s+1)(b+1)$.

Lemma 2. *Let $p_1^{m_1} p_2^{m_2} \cdot p_k^{m_k}$ be the prime factorization of $\frac{v}{4}$. Then $\sigma(v) = \prod_{i=1}^{k} \frac{p_i^{m_i+1} - 1}{p_i - 1}$.*

Proof. The set of pairs (s, b) that satisfy (i) $s \geq 0$. (ii) $b \geq 0$, and (iii) $v = 4(s+1)(b+1)$ is in one-to-one correspondence with the factors of $\frac{v}{4}$, by the correspondence that takes a factor d to the pair $(s, b) = \left(d - 1, \frac{v}{4d} - 1\right)$. For each such pair (s, b), corresponding to the factor $d = s+1$ of $\frac{v}{4}$, there are $s+1$ triples (s, b, f) that satisfy (iv) $0 \leq f \leq s$. Therefore, the number of triples (s, b, f) that satisfy (i), (ii), (iii), and (iv) is equal to the sum of the factors d of $\frac{v}{4}$. This sum is equal to $\prod_{i=1}^{k} \frac{p_i^{m_i+1} - 1}{p_i - 1}$. □

Proposition 7. *Let $p_1^{m_1} p_2^{m_2} \cdot p_k^{m_k}$ be the prime factorization of $\frac{v}{4}$. Then*

$$\frac{1}{3}\prod_{i=1}^{k}\frac{p_i^{m_i+1}-1}{p_i-1} \le \alpha(v) \le \prod_{i=1}^{k}\frac{p_i^{m_i+1}-1}{p_i-1}$$

and

$$\frac{1}{6}\prod_{i=1}^{k}\frac{p_i^{m_i+1}-1}{p_i-1} \le \beta(v) \le \prod_{i=1}^{k}\frac{p_i^{m_i+1}-1}{p_i-1}$$

Proof. If each of the triples (s, b, f) with $s \ge 0$, $b \ge 0$, $0 \le f \le s$, and $(s+1)(b+1) = \frac{v}{4}$ represented a distinct trihex, there would be $\sigma(v)$ distinct trihexes with v vertices. However, some of these triples represent equivalent trihexes, since a trihex can be decomposed in three ways into spines. Therefore $\frac{\sigma(v)}{3} \le \alpha(v) \le \sigma(v)$. There could be up to six signatures that represent isomorphic trihexes, since the left-handed and right-handed versions of chiral trihexes each have their own three signatures. So $\frac{\sigma(v)}{6} \le \beta(v) \le \sigma(v)$.

□

Table 2 gives the number $\alpha(v)$ of equivalence classes of trihexes and the number $\beta(v)$ of graph isomorphism classes of trihexes for each number v of vertices for $0 \le v \le 200$. Figure 18 presents the same information graphically for $0 \le v \le 400$. Note that the counts in the $\beta(v)$ column of Table 2 are always one greater than the counts in Table 5 of [1], since our counts include non-convex godseyes, and there is exactly one godseye for each possible number of vertices.

For each line in Table 2, the count $\alpha(v)$ is close to $\left\lceil \frac{\sigma(v)}{3} \right\rceil$ and $\beta(v)$ is fairly close to $\left\lceil \frac{\sigma(v)}{6} \right\rceil$. For $200 \le v \le 4000$, the difference $\alpha(v) - \left\lceil \frac{\sigma(v)}{3} \right\rceil$ is greater than one only 5% of the time and is never more than four. For $200 \le v \le 4000$, $\left(\frac{\sigma(v)}{3}\right) \le \alpha(v) \le 1.06\left(\frac{\sigma(v)}{3}\right)$. For $200 \le v \le 4000$, the difference $\beta(v) - \left\lceil \frac{\sigma(v)}{6} \right\rceil$ is greater than one 74% of the time and has a maximum value of 22. For $200 \le v \le 4000$, $\left(\frac{\sigma(v)}{6}\right) \le \beta(h) \le 1.25\left(\frac{\sigma(v)}{6}\right)$.

Conjecture 1. *As $v \to \infty$, the counts $\alpha(v)$ and $\beta(v)$ are respectively asymptotic to $\frac{1}{3}\sigma(v)$ and $\frac{1}{6}\sigma(v)$.*

v	$\alpha(v)$	$\lceil\frac{\sigma(v)}{3}\rceil$	$\beta(v)$	$\lceil\frac{\sigma(v)}{6}\rceil$	v	$\alpha(v)$	$\lceil\frac{\sigma(v)}{3}\rceil$	$\beta(v)$	$\lceil\frac{\sigma(v)}{6}\rceil$
4	1	1	1	1	104	14	14	8	7
8	1	1	1	1	108	14	14	9	7
12	2	2	2	1	112	20	19	13	10
16	3	3	3	2	116	10	10	6	5
20	2	2	2	1	120	24	24	14	12
24	4	4	3	2	124	12	11	7	6
28	4	3	3	2	128	21	21	15	11
32	5	5	5	3	132	16	16	10	8
36	5	5	4	3	136	18	18	10	9
40	6	6	4	3	140	16	16	10	8
44	4	4	3	2	144	31	31	20	16
48	10	10	8	5	148	14	13	8	7
52	6	5	4	3	152	20	20	11	10
56	8	8	5	4	156	20	19	12	10
60	8	8	6	4	160	30	30	20	15
64	11	11	9	6	164	14	14	8	7
68	6	6	4	3	168	32	32	18	16
72	13	13	8	7	172	16	15	9	8
76	8	7	5	4	176	28	28	17	14
80	14	14	10	7	180	26	26	16	13
84	12	11	8	6	184	24	24	13	12
88	12	12	7	6	188	16	16	9	8
92	8	8	5	4	192	42	42	28	21
96	20	20	15	10	196	21	19	12	10
100	11	11	7	6	200	31	31	17	16

Table 2: Count $\alpha(v)$ of equivalence classes of trihexes and count $\beta(v)$ of graph isomorphism classes of trihexes. Here $\sigma(v)$ is the sum of the factors of $\frac{v}{4}$.

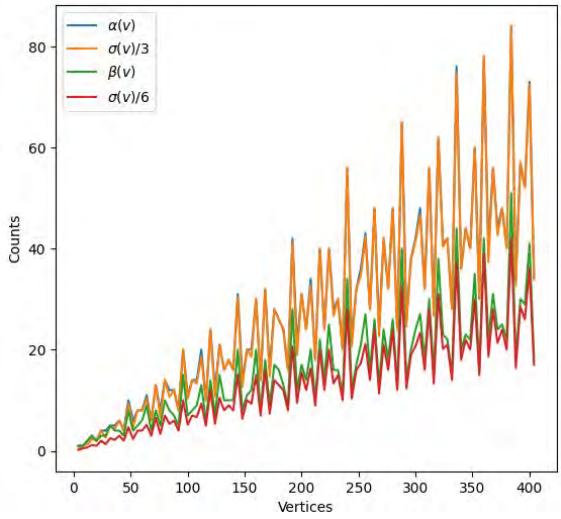


Fig. 18: The number of distinct trihexes that contain v vertices.

8 Tight Trihexes

This section analyzes trihexes that contain no belts. We start with a specialization of a definition from [1].

Definition 12. *A trihex is* tight *if it does not contain any belts.*

Proposition 8. *The following are equivalent for a trihex:*

1. *The trihex is tight.*
2. *The second coordinate in each of its three signatures is 0.*
3. *The trihex has a signature (s_0, b_0, f_0) such that the three numbers f_0, $f_0 + 1$, and $s_0 + 1$ are pairwise relatively prime and $b_0 = 0$.*
4. *For each of its three signatures, (s_i, b_i, f_i) for $i = 1, 2, 3$, f_i, $f_i + 1$, and $s_i + 1$ are pairwise relatively prime and $b_i = 0$.*

Proof. $(i) \iff (ii)$: Suppose the trihex is tight. Then the second coordinates in each of its three signatures must be zero, since the second coordinate indicates the number of belts between spines. Conversely, if the second coordinate in each of its three signatures is 0, then in the hexagonal tiling that covers the trihex, there are no infinite strips of hexagons that do not contain special hexagons, either in the vertical direction, the NW to SE direction, or the SW to NE direction. But any belt in the trihex would lift to such a strip, so the trihex must be tight.

$(ii) \implies (iv)$: Consider the three signatures (s_1, b_1, f_1), (s_2, b_2, f_2), and (s_3, b_3, f_3). Since the number of hexagons $h = 2s_ib_i + 2s_i + 2b_i$ for $i = 1, 2, 3$, and $b_i = 0$, we have that $s_1 = s_2 = s_3$. In the algorithm given in Definition 11, $s_2 = j_2(b_1 + 1) - 1$, where j_2 is the order of f_1 in Z_{s_1+1}. Since $b_1 = 0$ and $s_2 = s_1$, this means that $s_1 = j_2 - 1$, that is, f_1 has order $s_1 + 1$ in Z_{s_1+1}. Equivalently, f_1 is relatively prime to $s_1 + 1$. Similarly, $s_3 = j_3(b_1 + 1) - 1$, so $s_1 = s_3 = j_3 - 1$, where j_3 is the order of $f_1 + b_1 + 1$ in Z_{s_1+1}, that is, the order of $f_1 + 1$ in Z_{s_1+1}. So $f_1 + 1$ has order $s_1 + 1$ in Z_{s_1+1} and is therefore relatively prime to $s_1 + 1$. Since f_1 and $f_1 + 1$ are relatively prime to each other, we have that f_1, $f_1 + 1$, and $s_1 + 1$ are pairwise relatively prime. The same argument holds regardless of which of the three signatures we call (s_1, b_1, f_1). Therefore, f_i, $f_i + 1$, and $s_i + 1$ are relatively prime for $i = 1, 2, 3$.

Clearly $(iv) \implies (iii)$.

$(iii) \implies (ii)$: Assume without loss of generality that $b_1 = 0$ and f_1, $f_1 + 1$, and $s_1 + 1$ are relatively prime, since the signature (s_0, b_0, f_0) can be taken to be the first signature. Since $b_1 = 0$, the number of hexagons $h = 2s_1$. In Definition 11, since f_1 is relatively prime to $s_1 + 1$, $j_2 = s_1 + 1$. Therefore, $s_2 = j_2(b_1 + 1) - 1 = (s_1 + 1)(0 + 1) - 1 = s_1$, so $b_2 = \dfrac{h - 2s_2}{2s_2 + 2} = \dfrac{h - 2s_1}{2s_1 + 2} = b_1 = 0$.

Furthermore, since $f_1 + b_1 + 1 = f_1 + 1$ is relatively prime to $s_1 + 1$, $j_3 = s_1 + 1$. So $s_3 = j_3(b_1 + 1) - 1 = (s_1 + 1)(0 + 1) - 1 = s_1$, and $b_3 = \dfrac{h - 2s_3}{2s_3 + 2} = \dfrac{h - 2s_1}{2s_1 + 2} = b_1 = 0$. □

Remark 1. *Recall from Lemma 2 that the number of vertices in a trihex with signature (s, b, f) is $v = 4(s + 1)(b + 1)$. Therefore, Proposition 8 offers alternative proofs of Theorems 5.4 and 5.5 from [1], which can be restated as:*

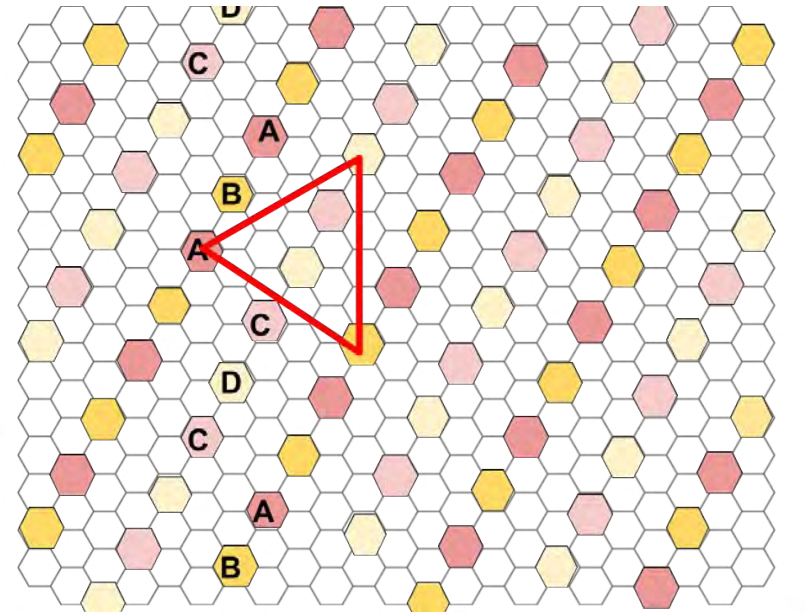


Fig. 19: Hexagonal cover for a tight trihex.

1. *If a trihex with v vertices is tight, then $\frac{v}{4}$ is odd.*
2. *For a number v, all the trihexes with v vertices that are not godseyes are tight if and only if $\frac{v}{4}$ is prime or equal to 1.*

Our next result proves Conjecture 5.3 from Deza and Dutour [1]. We first repeat their definition of the graph of curvatures:

Definition 13. *The* graph of curvatures *of a trihex is the graph whose vertices are the four triangular faces. Two vertices c and d are connected by an edge if there exists a pseudo-road connecting the faces c and d. A* pseudo-road *is a sequence of hexagons, say $a_1, \ldots, a_\ell$, such that setting $a_0 = c$ and $a_{\ell+1} = d$, we have that for $1 \leq i \leq \ell$, a_i,is adjacent to a_{i-1} and a_{i+1} on opposite edges.*

Proposition 9. *The graph of curvatures of any tight trihex is a complete graph on 4 vertices.*

Proof. Consider a tight trihex with signature (s_1, b_1, f_1). Recall that the trihex arises as the quotient sphere from a hexagonal tiling of the plane, where the "special hexagons", which correspond to the four triangles, lie on the vertices of a superimposed parallelogram grid. Label these special hexagons A, B, C, and D as in Figure 19, so that hexagons labeled A and C lie in vertical columns, and hexagons labeled B and D lie in alternate vertical columns.

By Proposition 8, $b_1 = 0$ and f_1 and $f_1 + 1$ are both relatively prime to $s_1 + 1$. Therefore, $s_1 + 1$ must be odd. In Definition 11, j_2 and j_3, which are the orders of f_1 and $f_1 + 1$ in Z_{s_1+1}, respectively, must both equal $s_1 + 1$ and therefore also be odd. Since the columns of hexagons alternate between columns containing A and C and

columns containing B and D, if we translate a special hexagon A in the SW to NE direction by $j_2 = s_1 + 1$ columns, it will hit a special hexagon in the B/D column. Similarly, if we translate A in by $j_3 = s_1 + 1$ columns in the NW to SE direction, it will hit a special hexagon in the B/D column. Form a triangle with vertices at the centers of the original special hexagon A and these two special hexagons. The sides with a vertex in A both have the same length and are at a 60° angle from each other. Therefore, this triangle must be an equilateral triangle. So the two special hexagons in the B/D column must be s_1 hexagons apart, so one of them is a hexagon B and the other a hexagon D. Therefore, the graph of curvatures for the trihex must include an edge from A to B and an edge from A to D. Of course, if we translate a special hexagon A straight north or south, it will hit a special hexagon C, so the graph of curvatures also includes an edge from A to C. By repeating this argument starting with hexagons B, C, and D, in turn, instead of A, we see that each vertex in the graph of curvatures connects to every other vertex. □

The classification of trihexes based on signatures, as presented in this paper, may also be useful in addressing Conjecture 5.7 from Deza and Dutour [1] concerning symmetry types of trihexes. The technique of lifting polyhedra to the Euclidean plane or to a hyperbolic plane may offer insights into the structures of polyhedra with other combinations of faces.

Acknowledgments

We would like to express our appreciation to our colleague Bob Proctor for excellent suggestions and edits to an early version of this manuscript.

Declarations

Funding Information

No funds, grants, or other support was received.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

Both authors contributed to the research and read and approved the final manuscript.

Data Availability

All data generated or analyzed during this study are included in this published article. Python code used to generate this data is available at https://github.com/stellenli802/Trihex-Research.

References

[1] Deza, M., Dutour, M.: Zigzag structures of simple two-faced polyhedra. Combinatorics, Probability and Computing **14**(1-2), 31–57 (2005) https://doi.org/10.1017/S0963548304006583

[2] Deza, M., Sikirić, M.D.: Geometry of Chemical Graphs: Polycycles and Two-faced Maps vol. 119. Cambridge University Press Cambridge, Cambridge (2008)

[3] Grünbaum, B., Motzkin, T.S.: The Number of Hexagons and the Simplicity of Geodesics on Certain Polyhedra. Canadian Journal of Mathematics **15**, 744–751 (1963) https://doi.org/10.4153/CJM-1963-071-3

[4] Brinkmann, G., Goedgebeur, J., McKay, B.D.: The generation of fullerenes. Journal of chemical information and modeling **52**(11), 2910–2918 (2012) https://doi.org/10.1021/ci3003107

[5] Grünbaum, B., Klee, V., Perles, M.A., Shephard, G.C.: Convex Polytopes vol. 16. Springer, New York (1967)

[6] Ziegler, G.M.: Lectures on Polytopes vol. 152. Springer, New York (2012)

[7] Whitney, H.: 2-isomorphic graphs. American Journal of Mathematics **55**(1), 245–254 (1933) https://doi.org/10.2307/2371127

[8] Mohar, B., Thomassen, C.: Graphs on Surfaces vol. 10, pp. 39–40. JHU press, Baltimore (2001)

Meet the Author:

double major in business and statistics and analytics. Physics is one of her extracurricular passions and she enjoys researching acoustics and optics. She is also interested in exploring statistical and computational technologies used in physics research, such as Vernier Logger Pro.

Author: Karina Samuel

Class: 2026

Independent Research

Title: Investigation of the Effects of the Dimensions of an Instrument and the Sound it Emits

Research Question: How well does the Helmholtz resonator equation predict the volume-frequency relationship in guitars?

Abstract:

This study delves into the relationship between the volume of guitar bodies and the frequency of sound produced, a topic less explored compared to wind instruments. Unlike wind instruments, guitars possess a constrained size range and are influenced by string properties, complicating the volume-frequency study. The research utilizes wooden boxes of varying volumes to simulate guitar bodies, with controlled factors such as soundhole area and soundboard thickness, aiming to understand how volume impacts resonant frequency in line with the Helmholtz resonator concept.

The hypothesis suggests larger guitar volumes yield lower frequencies, based on the Helmholtz resonator equation and existing literature on woodwind instruments. Birchwood box models were constructed, and resonant frequencies were measured through tapping, recorded with Vernier microphone and Loggerpro technologies.

Results indicate a significant deviation between experimental and theoretical values from the Helmholtz resonator equation, implying potential systematic errors in the simplified guitar model and its classification as a Helmholtz resonator. However, an inverse relationship between volume and frequency was consistently observed.

While the study supports the general trend predicted by the Helmholtz equation, it highlights the equation's limitations in accurately modeling guitars, likely due to their complex structural nature. Future research directions include exploring alternative equations and experimental approaches for more precise representation of guitar acoustics. This study advances the understanding of musical acoustics in string instruments and suggests new methodologies for investigating guitar resonance.

Key Words: Guitar Acoustics, Resonant Frequency, Helmholtz Resonator, Volume-Frequency Relationship, Experimental Methodology

Contents:

Scope of Work:

The relationship between the volume of an instrument body and pitch is a well-researched concept, particularly in instruments of the woodwind family. For example, in all wind instruments, the pitch is generally dictated by the length of the column used ("The Science of Making Musical Instruments"). Long-columned instruments, like the contrabassoon, are capable of playing the lowest-pitched notes in the orchestra, while piccolos, which are generally 32 cm in length or less, play the highest-pitched notes. In physics, pitch is the perception of a particular frequency (Klapuri). Lower pitched instruments like the contrabassoon emit low frequency pressure-time waves with large periods, while high pitched instruments emit high frequency pressure-time waves with small periods. This phenomenon can be explained by the mathematical relationship for the speed of a wave below:

$c=\lambda f$, where c is the speed of sound, lambda (λ) is the wavelength of the wave, and f is the frequency of the oscillation.

When a note is played on an instrument, the air molecules vibrate in a set of alternating compressions and rarefactions (areas of high and low pressures). This causes a disturbance which propagates throughout the instrument body. As the wave propagates, a standing wave is formed (see II. Background Information). Since speed is the product of wavelength and frequency, and wavelength increases proportionally to the length of the instrument, frequency must decrease correspondingly (Nehru).

This investigation attempts to examine the volume and frequency relationship in guitars. Modern guitars also come in various sizes, similar to woodwind instruments. However, they do not have as significant of a range as that of other instrument families. Because of this as well as several external variables like mass, length, and tension of guitar strings, and each of their respective effects on frequency, it is much more challenging to study the guitar volume-frequency relationship than that of other instrument families.

Background Information:

a. <u>The Guitar</u>

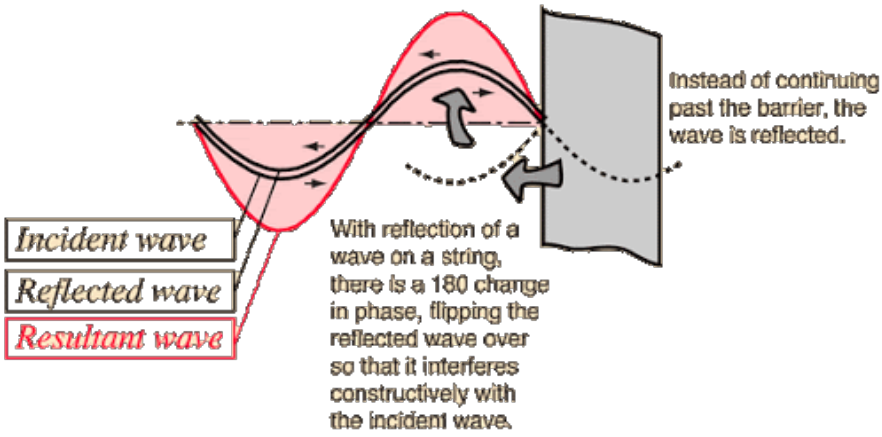
Guitars originated in Spain in the early 16th century and are derived from the Spanish Vihuela of the lute family of instruments. Spanish musician Antonio de Torres Jurado's 18th-century style of guitar craftsmanship gave rise to the shape and curvature of the modern guitar ("Guitar History: How the Guitar Has Evolved"). However, these 18th-19th century instruments were much smaller in size compared to modern day instruments. Martin dreadnought guitars, or "D-size" guitars, have since become the most common size for acoustic guitars. A typical acoustic guitar of this size is around 96.5 cm in length.

b. <u>Standing Waves</u>

Standing waves are waves that occur within a medium and share a consistent amplitude. All points in a standing wave move coherently with a constant phase difference and a shared frequency. Standing waves are created by the

superposition of the incident and reflected waves. This phenomenon is depicted in *figure 1* below. Standing waves have nodes (areas of minimum amplitude and maximum destructive interference) and antinodes (areas of maximum amplitude and maximum constructive interference).

Figure 1:



Source: "Image of Standing Wave Caused by Interference ." *Standing Waves* , Hyperphysics

Guitar strings generally vibrate in a number of different standing wave frequencies; these frequencies are known as *resonant frequencies* ("Standing Waves"). The resonant frequency is achieved when the object oscillates at its natural frequency. In other words, the object's frequency is equivalent to the frequency the object would have in the absence of any driving force. The fundamental frequencies of an in-tune guitar range from 80 hz to 1200 hz (Case).

c. Helmholtz Resonator

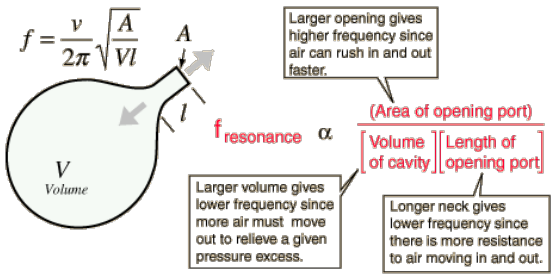
A Helmholtz resonator is defined as a container of gas with an open soundhole ("Helmholtz Resonance"). The soundhole and the volume of air enclosed within a guitar body can be compared to the Helmholtz resonator. As air oscillates through the soundhole, the volume of the body of the instrument dictates the frequency ("The Acoustic Guitar Body-Part 2"). This relationship can be modelled by the following formula:

$$f = v/2\pi(\sqrt{A/VL})$$

The variables specific to studying guitar frequency are as follows: *f* is the resonant frequency of the Helmholtz resonator, *v* is the speed of sound in air (roughly 343 meters per second), *A* is the area of the soundhole, *V* is the volume of air in the guitar chamber, and *L* is thickness of the soundboard.

The diagram below models this equation.

Figure 2:



Source: *Overview of Resonant Systems* , Hyperphysics

Theoretically, the larger the volume V the lower frequency f; the frequency-volume ratio of a Helmholtz resonator is an inversely proportional, square-root relationship. Mathematically, $f \propto \sqrt{1/V}$.

Methods:

In order to investigate this subject, an experiment will be conducted modelling the structure of a guitar. Laser-cut wooden boxes within a range of volumes will be used to represent the varying volumes of a guitar. Birchwood was chosen to best replicate the acoustic properties and traditional resonances of acoustic guitars. Soundholes will be cut from these boxes and will remain a controlled and constant area to ensure that volume is the only manipulated variable of the Helmholtz resonator Equation. Because the guitars are modeled by rectangular prisms, the volume V can easily be calculated and manipulated using $V=lwh$, where l, w, and h are length, width, and height respectively. All variables with the exception of the volume (the independent variable) and the frequency (the dependent variable) will be controlled, including the speed of sound, the area of the soundhole, and the thickness of the soundboard (the thickness of the birchwood sheets used to build the box).

Resonant frequency can be created by tapping each of the boxes (making the box “sing” a “note” or a particular frequency). This is the preferred means of producing sound, as using a string can introduce extraneous variables like tension and length [of the string] into the experiment. A vernier microphone and Loggerpro technologies will be utilized to measure the raw data frequency of the note played on each of the model guitars. After the initial investigation, the theoretical data collected using the Helmholtz resonator

equation will be compared to the experimental frequencies of the notes played by the boxes.

Investigation:

a. Hypothesis:

There are two main reasons why I believe that larger volumes will equate to lower frequencies. First, when examining the Helmholtz resonator equation itself, it is clear that increasing the volume will decrease the frequency. This is because mathematically, increasing the denominator value of the inverse relationship will lead to values closer to 0 (lower frequencies would result from larger volumes or soundboard thicknesses because these variables are in the denominator of the Helmholtz resonator Equation). If a guitar can truly be modelled by the Helmholtz resonator equation, this should hold true in my experiment.

The second reason I believe that larger guitar volumes will lead to lower frequencies is because of my background research. After reviewing some existing literature on length-frequency relationships in woodwind instruments, it seems as though instruments with more considerable dimensions will produce lower frequencies.

b. Variables:

The independent variable for this experiment is the Volume of the Guitar Model *(V)*. This variable is measured by the formula for a rectangular prism, $V=lwh$, where l is length, w is width, and h is height. The frequency of the box, f, is the dependent variable. The fundamental frequency will be measured using a Vernier Microphone and Loggerpro data plots relating time to sound

pressure, as well as an FFT graph relating frequency to amplitude. The thickness of the "soundboard" (the thickness of the box), *L*, is a controlled variable. The birchwood sheets that are used for the boxes are 3 millimeters thick across all the guitar models. The speed of sound, *v*, is another controlled variable, at roughly 343 meters per second. Lastly, the Area of the Soundhole, *A*, is controlled, at 45.96 square centimeters. The area of the soundhole is derived from the average diameter of a guitar soundhole, 7.65 cm. Using the formula $A=\pi r^2$, the area of the soundhole must be 45.96 cm^2 for all box numbers.

c. Materials:

While the complete materials list is detailed in Appendix B, the most important materials required for this experiment include a laser cutter (I used an Epilog Helix Mini Laser Machine at 75 watts), birchwood sheets and glue for cutting and assembling the apparatuses, a vernier microphone and appropriate adapters, and a computer with sufficient capabilities to run Loggerpro and laser cutting technologies (I used Adobe Illustrator). Dimensions of the boxes are listed in detail in Appendix A. The ratio between Height, Length, and Width of the boxes is 1:2:4 for all boxes.

I selected rectangular prisms to represent the guitar volumes over other shapes for two reasons. First, they are the easiest figure to create using the laser cutters, as it only had to perform straight line cuts outside of the circle in the center. It is also

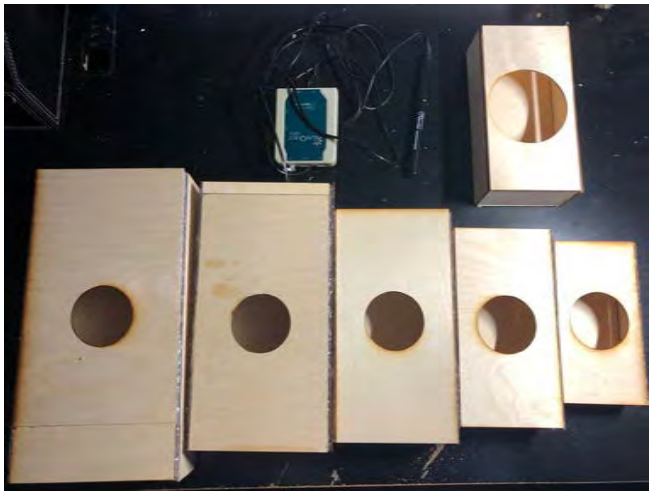
easy to complete volume calculations with box shaped rectangular prisms because the formula is very straightforward. Although it is easy to use this specific structure, the boxes are not a true representation of the shape of a guitar. This could potentially create some inaccurate data, because a rectangular prism is likely not the most accurate model.

I decided to use the laser cutter rather than cutting my boxes by hand because the laser cutter is much more precise. Moreover, when I did try hand cutting the wood using a saw, it left sharp frayed edges that could have potentially caused splinters. It was also very difficult to cut through the wood in a circular shape, making the process overly laborious. Thus, the laser cutter was the most efficient way to create my apparatus. I used birchwood sheets because they are often used to construct guitars and they are readily available and inexpensive. This was especially important in my investigation because I had to experiment several times with the laser cutter and burn through many sheets to figure out the most efficient procedure.

Furthermore, due to limits of the size of the laser cutter bed, I could not study boxes of volumes above 5832 cm^2, which is quite small compared to the size of the average D-size guitar. Below is a picture of the boxes I used placed in descending order of volume. Above the boxes on the left is the Vernier Microphone I used to record the data. To the right of the microphone is a prototype box that I built before creating the final apparatuses. This prototype box

contains a larger soundhole and contains large gaps in between pieces.

Figure 3:



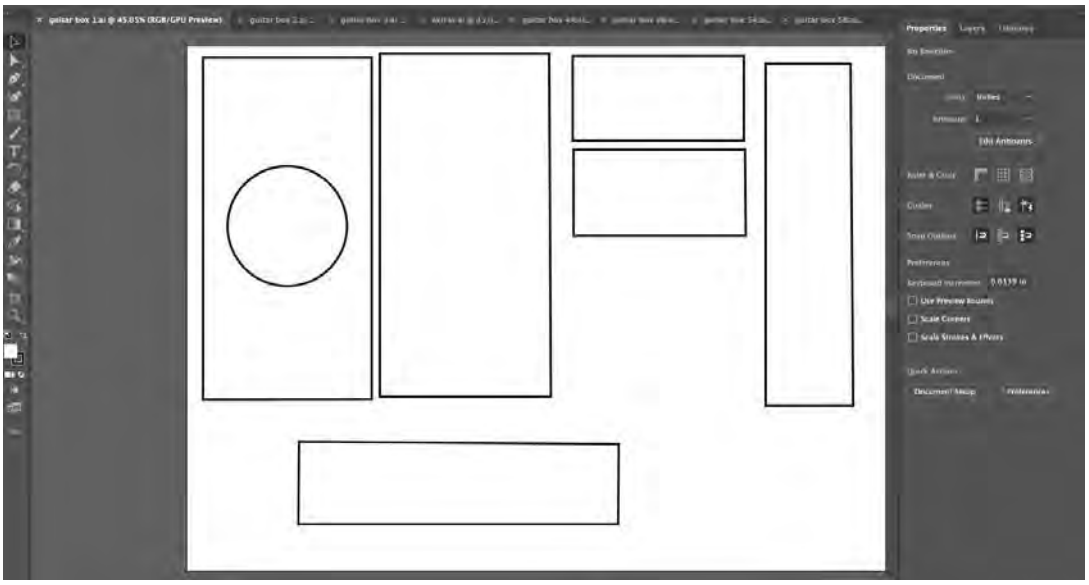
d. Procedure:

To study the relationship between the volume of the guitar models and their frequency, I began with cutting the boxes using the laser cutters. Refer to Appendix C to see the complete procedure with thorough steps detailed.

For my final boxes that I ended up testing in my experiment, I made the soundhole smaller to more accurately reflect the average size of a guitar soundhole. Moreover, I used hot glue to seal up the edges of each box. My prototype box also had decimal dimensions as I did not convert from centimeters to inches prior to inputting values into the laser cutter software. I made sure to make this conversion for my final apparatuses so I could work with cleaner whole numbers and ratios.

Below is an image of the *Adobe Illustrator* File used in this experiment. This application allows the user to delineate all six faces of the rectangular prism with the appropriate dimensions. Because the Epilog Laser Machine does not take SI metric unit inputs, I made sure to convert centimeters to US Customary Units (inches) before putting values into the software.

Figure 4:



When gluing, I used super glue to attach the pieces together and hot glue to add an extra seal to the box. I decided to do this because I am modelling these boxes as Helmholtz resonators, which must be sealed in order to accurately be represented by the Helmholtz resonator equation. When I set up the vernier microphone on Loggerpro, I originally saw that the software only related time to Sound Pressure and I could not see the amplitude of the wave or the frequency. To fix this, I had to use the Fast Fourier Transform (FFT) diagram that relates

frequency and amplitude. To create the resonance, I simply tapped the box with my finger near the top of the box. I recorded five trials for each box volume and took the average to find the mean resonant frequency each guitar model produced.

e. Safety:

The only major safety concerns for this experiment are the use of the laser cutter and the hot glue. However, any risks associated with laser cutters can be easily prevented with a few safety precautions. Moreover, because most laser cutter systems will not operate unless closed and sealed off, there is an especially low risk of fire.

1. While operating the laser cutter, I made sure to follow all the recommended safety precautions designated by the manufacturer. For the Epilog Helix Mini Laser Machine, these included:
    1. Turn on the laser cutter exhaust system. This will prevent inhalation of toxic fumes and fine particles.
    2. Ensure that the laser cutters have fully functioning covers and interlocks.
    3. Ensure that you do not remove the wood until the cutter bed has cooled.
    4. Clean up all debris and potentially flammable materials after cutting.

5. Keep a fire extinguisher nearby.

2. While operating the glue gun, I followed the following safety precautions. These were important in my experiment because I had to use a large quantity of hot glue to seal up my boxes, particularly the boxes with larger dimensions.
    1. Keep the hot glue gun away from flammable materials
    2. Place the glue gun on a safety stand when not in use
    3. Do not touch the nozzle
    4. Unplug the glue gun after use

Data Analysis:

a. Calculations:

To calculate the theoretical frequency using the Helmholtz resonator Equation, I had to convert all my values to correct SI base units. This involved converting the Area, A into meters squared, the Volume, V into meters cubed, and the Length, L, into metres.

I conducted my experiment and completed the calculations for experimental and theoretical values for each box. For the experimental data, I took the average of the five trials. For the theoretical data, I completed the calculation from the Helmholtz resonator equation to solve for f for each box.

b. Processing Raw Data:

Absolute Error: Absolute error can be calculated for each box using the formula, (Maximum - Minimum)/Number of data points.

Example Calculation (box 5):

[(217 hz - 199 hz)/5]=3.6

Percent Error (error between readings): Percent error between experimental readings can be calculated using the formula, (Absolute Error/Average) × 100%.

Example Calculation (box 5):

(3.6/205) x *100% =1.76%*

Percent Error (error between experimental and theoretical values): The error between the experimental frequencies and the accepted frequencies calculated by the Helmholtz resonator formula can be found using the formula, [(experimental value-theoretical value)/theoretical value] x 100%.

Example Calculation (box 5):

[|(205 hz -846 hz)/846 hz|] x *100% =75.8%*

Below is the table of processed values with error.

Table 1:

BOX	Volume (cm^3)	Measured		Calculated from Helmholtz resonator Equation	
		Average Frequency (hz)	% Error	Frequency (hz)	% Error (error between theory and experimental values)
1	1000	535	.82	2136	74.8

2	1728	412	.87	1625	74.9
3	2744	304	.97	1289	76.3
4	4096	273	.95	1056	74.4
5	5832	205	1.76	846	75.8

c. Data Linearization:

To linearize the Helmholtz resonator Equation, I plotted $\sqrt{1/volume}$ on the x-axis and frequency on the y-axis. The data table used to create the $\sqrt{1/V}$ versus frequency graph (see *figure 5)* is below. The error bars show the error between individual frequency readings during the five trials. Refer to Appendix E to see unprocessed data.

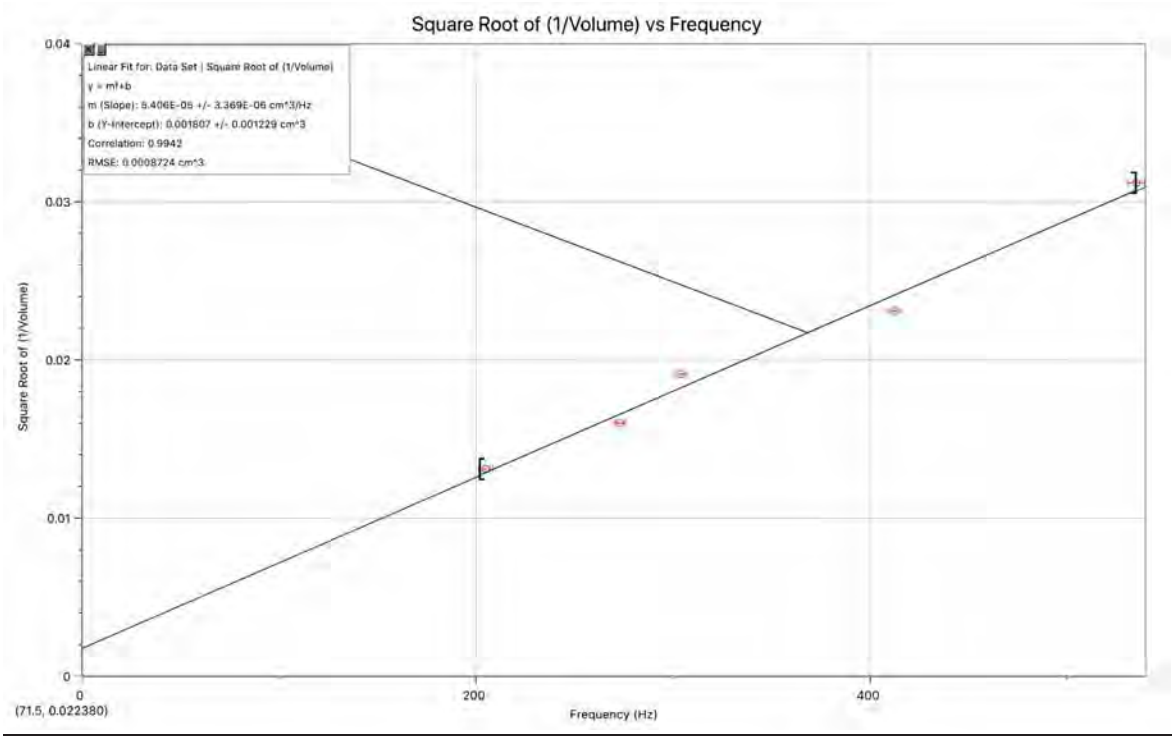
Data table for the resonant frequency versus the square root of the inverse of the volume with percent error of frequency:

Table 2:

Frequency (hz)	$\sqrt{1/volume}$(cm^3)	Error (of frequency)
535	.0312	.82
412	.0231	.87
304	.0191	.97
273	.016	.95
205	.0131	1.76

d. Presenting Processed Data:

Figure 5:



e. Qualitative Comments

I found that there is a very clear systematic error in the data--the percentage error between the experimental and theoretical values range from 74.4% to 76.3%. This consistent and significant difference between the values indicates that there might be a flaw in the construction of the apparatus. Moreover, before the absolute value of the percent error was taken, all the errors were negative. This indicates that the experimental or calculated values are significantly less than the theoretical or accepted values.

While the percent error between the experimental and theoretical frequencies is very high, the percent error between individual frequencies in the experimental data is very low (all below 2%). This indicates that the data is very precise and consistent but also very

inaccurate when compared to the theory.

In the graph above (*figure 5),* there is a clear positive linear relationship between the Frequency of the sound and the square root of the inverse of the volume. As Frequency increases, the square root of the inverse of volume also increases. The high correlation coefficient of .9942 indicates that the linear fit is a good predictive model for this relationship. This verifies the inverse relationship that as the volume of the guitar model increases, the frequency decreases.

Conclusion and Evaluation:

a. Conclusion:

After plotting frequency versus the square root of the inverse of the volume, I found that the inverse of the volume increased as frequency increased. In other words, higher volumes correlate with lower frequencies. It is important to note that the linearized data only presents a trend that, while directly proportional, does not go through the origin of the graph. This indicates that my experiment does have error. However, the trend supports that the general relationship given by the Helmholtz resonator Equation is somewhat accurate. This also supports my hypothesis that as the volume of the guitar model increases, the frequency decreases.

There is very little error between the individual frequency readings across the five trials. This suggests that the resonant frequency stayed generally the same across all trials, indicating that the data is very consistent and precise. However, I found that there was a very

significant amount of error when comparing the experimental data from my experiment and the theoretical values calculated from the Helmholtz resonator equation. This indicates that my data was very far off from the resonant frequencies that generally result from testing the volumes I selected.

After reviewing my data, I noticed an interesting relationship between the experimental and theoretical values of resonant frequencies: my model guitars' measurements were off by about a factor of 4. In other words, the theoretical values were roughly four times higher than my experimental data. For example, for box number 5, I calculated a resonant frequency of 846 hz from the Helmholtz resonator equation. However, my experimental data for box five averaged to about 207 hz. In this case, the theoretical frequency is slightly above four times the experimental value; this is consistent with the theoretical and experimental values for the four other boxes in my experiment as well.

b. Sources of Random Error:

Random error in the data could have arisen from two places. First, because the Epilog Helix Mini Laser Machine that was used for this experiment only took Inches as input values, a conversion between centimeters and inches was necessary. Since the conversion between centimeters and inches involves a division by a decimal factor (x centimeters=x/2.54 inches), I had to round the inch values for laser cutter input. This could have caused slightly inaccurate volume measurements for each box. Second,

the seal of the box could have been incomplete. Despite my efforts to reseal the box after the hot glue dried, it is possible that smaller or unnoticeable gaps could have opened up, compromising the box's Helmholtz resonator Quality (Helmholtz resonators must be completely sealed with the exception of the soundhole).

c. Sources of Systematic Error:

Despite the fact that my data somewhat supported the inverse square root relationship provided by the Helmholtz resonator Equation, it clearly does not accurately match up with the expected theoretical values. This is probably mostly attributable to large amounts of systematic error. There are a few areas in my experiment that could have contributed to systematic error. First, I did not model the guitars in the classic dreadnought shape (the curved structure); I instead used rectangular prisms to make my set up easier and to prevent my volume calculations from becoming too exhaustive. This could have led to systematic error because my apparatus was too poorly designed to accurately reflect what a true guitar's resonant frequency would be.

More important, however, is my classification of a guitar as a Helmholtz resonator at the beginning of my research. While a guitar can be modelled as or compared to a Helmholtz resonator, it is not a true representation of one (see II. Background Information). This can be clearly seen when the structures of a guitar (as well as the structure of my rectangular prism models) are compared to the classic structure of a Helmholtz resonator (refer to *figure*

2 in II. Background Information). In this picture, it is very obvious that a classical Helmholtz resonator is a circular, jug-like container with a long neck. This specific structure causes the air to compress within the container, increasing the pressure and thereby causing the air to vibrate in a particular way. My model guitars do not contain this long neck or curved structure; this could have contributed to my inaccurate results.

d. Evaluation of Procedures:

Evaluation of Online Sources:

The information I used for my background information as well as to find the Helmholtz resonator equation come from a broad array of scientific sources, many of which come from university professionals specializing in research regarding musical acoustics. For example, the article entitled "Helmholtz Resonance," which I cite in my paper several times, comes from a collection of publications in voice and music acoustics from The University of South Wales' research staff.

Evaluation of Methods and Improving the Investigation:

While my methods were somewhat effective in showing the relationship between the volume of an apparatus and its resonant frequency, it is not very effective at modeling a guitar as a Helmholtz resonator. To improve this investigation, I could test several alternative methods. For example, I could blow over the top of the guitar soundhole instead of tapping to create resonance, and then record the data. This could improve the

investigation because the Helmholtz resonator equation is derived from the change in pressure caused by an air jet over the neck of the container, rather than a tap ("Helmholtz Resonance"). To make my experiment more representative of guitar frequencies, I could study the driving frequencies of real guitars of different sizes rather than the resonant frequencies of boxes.

To improve my experiment without completely changing the apparatuses, I could utilize a wider range of guitar model volumes. In my experiment, I was only able to study the relationship across five boxes. While the range between the largest and smallest box is quite large (4832 cm^3, refer to Appendix A to see dimensions), using even larger volumes could more accurately represent the volume of a full sized guitar. This would also provide me with a larger sample size for the experiment, allowing me to plot more values and prevent large amounts of statistical extrapolation in my linearized graph (see *figure 5).*

However, I believe that my experiment is very easily repeatable and will likely yield a similar mathematical relationship if conducted again using the same methodology.

e. <u>Suggestions for Further Research:</u>

After exploring the volume versus resonant frequency relationship in a Helmholtz resonator, I researched some alternative equations that could better represent the mathematical relationship within a guitar. This equation is defined below by the article entitled "Helmholtz

Resonance": $f = (c/2\pi)(\sqrt{(\pi r^2)/(V \cdot 1.7r)})$, where *f* is the resonant frequency of a guitar, *c* is the speed of sound at 340 meters per second, *V* is the volume of the guitar, and *r* is the radius of the soundhole.

This research explains that the traditional Helmholtz resonator equation often overestimates the resonant frequency (exactly how the equation overestimated the resonant frequency in my experiment) due to "swelling" of the instrument body when air flows through it. The researchers suggest that in order to make the equation accurate, the body of the instrument must be kept constant throughout the experimentation by "burying the guitar in sand to impede the swelling or 'breathing' of the body" ("Helmholtz Resonance"). While this research does not go into detail about the derivation of this new and guitar specific Helmholtz resonator equation, it is possible that researching and possibly experimenting with this new equation could yield more accurate results than the traditional equation that I used for my experiment.

Appendix A: Table of Laser Cutter Inputs

Table 3:

Box Number:	Length (cm):	Width (cm):	Height (cm):	Area of Soundhole (cm^2)	Volume (cm^3)
1	20	10	5	45.96	1000
2	24	12	6	45.96	1728

3	28	14	7	45.96	2744
4	32	16	8	45.96	4096
5	36	18	9	45.96	5832

Appendix B: Full Materials List

1. 1 Laser Cutter (an Epilog Helix Mini Laser Machine at 75 watts was used for this experiment)
2. Birchwood Sheets of Appropriate Dimensions
    a. The area of the soundhole is derived from the average diameter of a guitar soundhole, 7.65 cm. Using the formula $A=\pi r^2$, the area of the soundhole must be 45.96 cm^2 for all box numbers. This is because the variable *A* (see 'Variables') must be controlled in order to calculate the frequency using the Helmholtz resonator Formula
    b. Table of Input Values for Laser Cutter (the ratio of Height:Length:Width is 1:2:4 for all boxes):

Table 4:

Box Number:	Length (cm):	Width (cm):	Height (cm):	Area of Soundhole (cm^2)	Volume (cm^3)
1	20	10	5	45.96	1000
2	24	12	6	45.96	1728
3	28	14	7	45.96	2744
4	32	16	8	45.96	4096
5	36	18	9	45.96	5832

3. 1 bottle of Super Glue

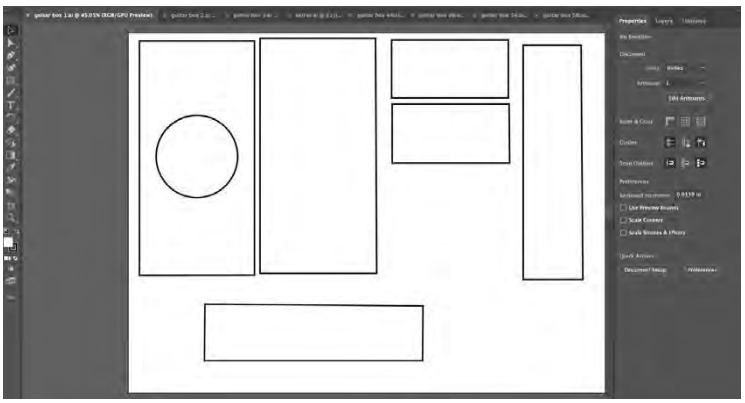
4. 1 Hot Glue Gun with refillable sticks
5. 1 Vernier Microphone
6. 1 Microphone to USB Adapter
    a. If necessary: 1 USB to USB-C adapter or equivalent (if computer does not use USB)
7. 1 Computer with sufficient software capabilities to open the *Loggerpro* application as well as any laser cutting software *(Adobe Illustrator, Inkscape, SketchUp, SolidWorks,* etc*). Adobe Illustrator* was used for this experiment.

Appendix C: Full Procedure

1. Use the laser cutter to cut boxes of the following dimensions from *table 1* above.

Below is an image of the *Adobe Illustrator* File used in this experiment. This application allows the user to delineate all six faces of the rectangular prism with the appropriate dimensions. Because the Epilog Laser Machine does not take SI metric unit inputs, be sure to convert centimeters to US Customary Units (inches) before putting values into the software.

Figure 3:



2. Glue all the boxes together using super glue. Ensure the containers are as airtight as possible.
3. Go over the edges of each of the boxes using hot glue to maximize the seal of the box. Ensure that when the glue dries, there are still no gaps between the wood pieces. Reseal if necessary.

4. Select Box 1. Attach the Vernier Microphone to a Microphone to USB adapter (or adapter equivalent) and plug into a computer with Loggerpro graphing capabilities. Open the Loggerpro application.
5. Set up the microphone on the Loggerpro application by clicking “Experiment” at the top left of the screen. Select the “set up sensors” option, and choose “LabQuest _1.”
    a. If necessary, change “Sound Level Sensor” to “Microphone” by clicking the “Choose Sensor” option. After switching this option, the graph should now relate time to Sound Pressure.
6. Select the clock button in the top row of options. Change the duration to 5 seconds. This will ensure that you have enough time to press record and play the “note” on the box.
7. Press record (the green button at the top) and create resonance by tapping the box in the area marked below.

Figure 7:

8. Select the area of the graph containing the “note” from the tap by highlighting that portion of the graph.

9. In order to see the frequency, access the Fast Fourier Transformation (FFT). Select “Insert” and “Additional Graphs,” and choose the FFT graph. The new graph that appears should relate Frequency and Amplitude.
10. In order to see the peak frequency, double click the graph and choose “Legend.” Turn on “peak frequency.” Now, the graph should show the range of values between which the peak frequency falls.
11. Record data and convert to FFT at least 5 times
12. Repeat steps 6-11 with boxes 2, 3, 4, and 5.
13. After you have all the raw data, calculate the average frequency for each box out of the three trials (add up the values for all three frequencies and divide by three).

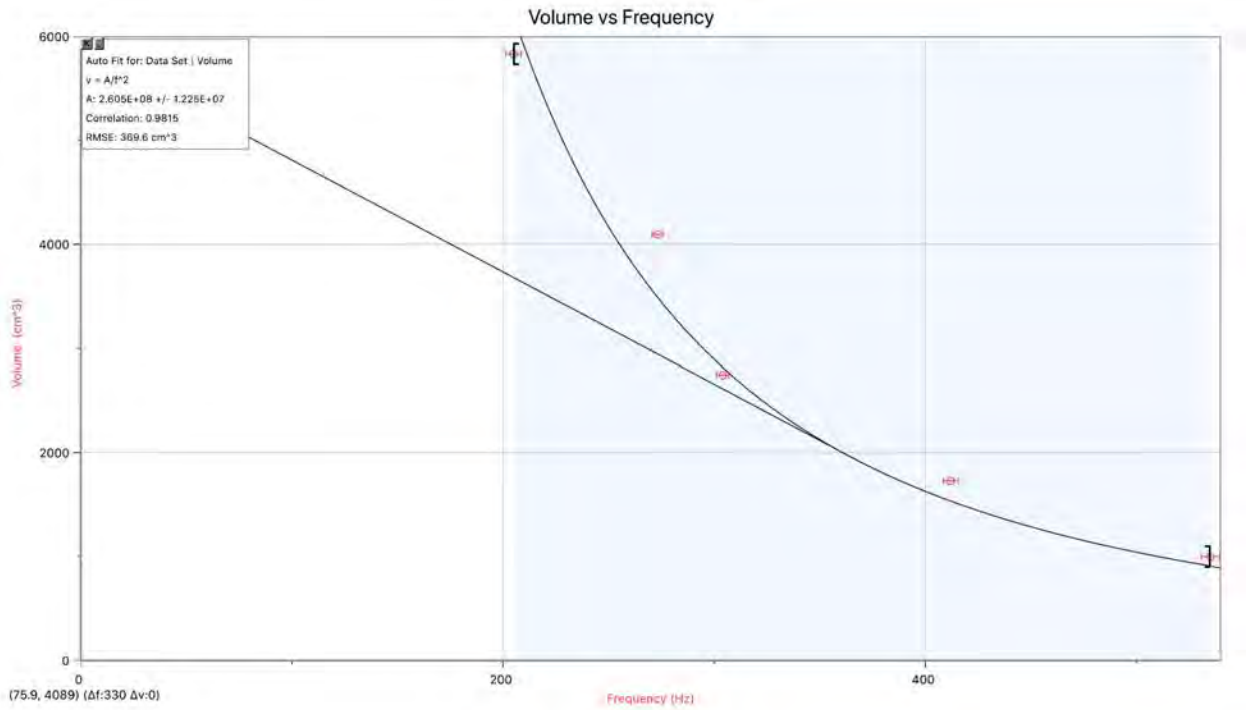
Appendix D: Raw Data

Box 1:	Frequency (hz):		**Box 2:**	Frequency (hz):		**Box 3:**	Frequency (hz):
Trial 1:	527		Trial 1:	407		Trial 1:	305
Trial 2:	549		Trial 2:	405		Trial 2:	312
Trial 3:	538		Trial 3:	412		Trial 3:	297
Trial 4:	528		Trial 4:	415		Trial 4:	302
Trial 5:	532		Trial 5:	423		Trial 5:	305
Average:	535		Average:	412		Average:	304
Calculated (from Helmholtz resonator formula):	2136		Calculated (from Helmholtz resonator formula):	1625		Calculated (from Helmholtz resonator formula):	1289
Percent Error (error between experimental data and theoretical	74.80%		Percent Error (error between experimental data and theoretical	74.90%		Percent Error (error between experimental data and theoretical	76.30%

data):			data):			data):	
Percent Error (error between readings):	0.82%		Percent Error (error between readings):	0.87%		Percent Error (error between readings):	0.97%
Box 4:	Frequency (hz):		**Box 5:**	Frequency (hz):			
Trial 1:	272		Trial 1:	217			
Trial 2:	271		Trial 2:	200			
Trial 3:	268		Trial 3:	199			
Trial 4:	272		Trial 4:	201			
Trial 5:	281		Trial 5:	207			
Average:	273		Average:	205			
Calculated (from Helmholtz resonator	1056		Calculated (from Helmholtz resonator	846			

formula):			formula):				
Percent Error (error between experimental data and theoretical data):	74.40%		Percent Error (error between experimental data and theoretical data):	75.80%			
Percent Error (error between readings):	0.95%		Percent Error (error between readings):	1.76%			

Appendix E: Unprocessed/Non linearized data



Works Cited

Anssi Klapuri, "Introduction to Music Transcription", in Signal Processing Methods for Music Transcription, edited by Anssi Klapuri and Manuel Davy, 1–20 (New York: Springer, 2006): p. 8. ISBN 978-0-387-30667-4.

Case, Alex. “Guitars.” *Recordingology: The Study of Recording* , recordingology.com/in-the-studio/guitars/#:~:text=The%20fundamental%20frequencies%20in%20the,at%20multiples%20of%20these%20frequencies.

“Guitar History: How the Guitar Has Evolved.” *Musicians Institute Hollywood*, 20 Feb. 2021, www.mi.edu/education/guitar-history-how-the-guitar-has-evolved/.

“Helmholtz Resonance”, The University of New South Wales , newt.phys.unsw.edu.au/jw/Helmholtz.html.

Nehru . “Musical Pipes .” *Taralaya*, www.taralaya.org/musical-pipes.html#:~:text=Musical%20Pipes&text=The%20pitch%20produced%20depends%20on,outer%20radii%2C%20but%20different%20lengths.

“Standing Waves .” *Salford Acoustics* , salfordacoustics.co.uk/sound-waves/standing-waves#:~:text=Because%20the%20standing%20waves%20occur,fundamental%20frequency%20(3F%20Hz).

“The Acoustic Guitar Body – Part 2.” *Acoustic Masters - Guitar Body 2*, www.acousticmasters.com/AcousticMasters_GuitarBody2.htm.

“The Science of Making Musical Instruments.” *OpenLearn*, The Open University, 30 Aug. 2019, www.open.edu/openlearn/science-maths-technology/science/physics-and-astronomy/physics/the-science-making-musical-instruments.

Meet the Author:

Hi! I'm Emily, I'm a senior majoring in Mathematics, with a minor in data science. In my spare time, I enjoy baking, making jewelry, and spending time with my dog. While I take a lot of math classes, I also like to mix things up with humanities classes, and this is a paper I wrote for one of AMST 276: Food and American Culture where I dive into the history of eggs.

Cockfighting to Capitalism:

The Historical Rise of the American Egg Industry

Emily Huang

AMST 276: Food and American Culture

February 20, 2023

From humble beginnings to a staple of the American diet, eggs have a long and complex history. Humans have been fascinated by eggs for centuries with the famous chicken and egg paradox inspiring philosophers to delve deep into the history of chickens. The history of chickens and eggs dates back thousands of years, before evolving into the modern egg industry we know today. Technological advancements and changes in consumer preferences have shaped the industry, leading to a shift towards more centralized and commercialized methods of farming, but also giving rise to concerns about animal welfare. Despite these criticisms, eggs remain a beloved and versatile food in American cuisine, and their cultural significance is deeply ingrained in our society.

The domestication of chickens for egg production dates back thousands of years. Scientists commonly agree that chickens originate from the Red Junglefowl and were domesticated around seven thousand to eight thousand years ago in Southeast Asia and Oceana (Zaheer 2015, 1208-1209). Through DNA studies with modern chickens, scientists have discovered separate domestications branching off throughout China and other parts of Asia thought to be bred for ritualistic purposes. Fighting cocks were bred in India and Southeast Asia and rapidly spread throughout the globe thanks to the accessibility of chickens and low cost of maintenance (Davis 2013, 552). Cockfighting is believed to be the primary reason that chickens were first distributed and used for food, a purpose we still see in the modern egg industry.

Throughout the nineteenth century, the chicken industry generally consisted of small family flocks that supplied most of their eggs for a single family's consumption. Small-scale family farms were the norm.

With the advancement of technology at the turn of the century, the egg industry began shifting towards a more centralized and commercialized model. Integrating new information, small-scale farmers began to move their hens indoors, which protected them from inclement weather, predators, and disease. However, there were still apparent problems regarding sanitation within the coop. With advancement in transportation and the discovery of sanitation benefits with a raised wire-floor cage, the modern egg industry truly began to take shape. In the 1950s, the battery cage system, which has wire floors that immediately remove manure and provide a controlled environment for chickens, was introduced and adopted by commercial egg farmers (Braunschweig-Norris 2006, 515). Large commercial farms emerged using these wire cages, and could supply eggs to urban areas and beyond. These commercial farms could maintain thousands of chickens and used new technologies and practices to increase production and lower costs.

The egg industry continued to evolve, shaped by technological advancements and changes in consumer preferences. Barren raised wire cages are cost-effective and efficient methods of farming eggs and continue to be the most prevalent hen housing method in the United States. “Laying hens are typically afforded forty-eight square inches per hen in the battery cage” and due to this limited space, the hens cannot stand properly or perform other natural behaviors (Braunschweig-Norris 2006, 516-517). The government has also played a significant role in the growth of the American egg industry. The New Deal, implemented by President Franklin D. Roosevelt, introduced programs that supported the modern egg industry, including the Rural Electrification Administration (REA) and the Farm Security Administration (FSA). These

programs brought electricity and financial aid to farmers, which in turn helped increase the efficiency of egg production (Rasmussen 1983, 1160). The increased efficiency also made eggs more affordable, thereby increasing demand.

Despite the successes of the American egg industry, there have been criticisms of certain production methods. Concerns have been raised about the welfare of the chickens and risks of consuming eggs produced in unideal conditions. The treatment of chickens in the egg industry does not meet the physiological and behavioral requirements for hens, which leads to a poor quality of life. Abnormal behaviors are exhibited in egg-laying hens living in battery cages, such as feather-pecking, cannibalism, and aggression (Shields and Duncan 2009, 4-5). Studies done suggest that the use of battery cages poses health risks to chickens and humans alike. The lack of exercise contributes to bone weakness and increased risk of injuries and deformities despite selective breeding to improve bone strength. As they are raised off of the ground, cages minimize exposure to soil-borne diseases and infectious agents, but , fresh air, sunlight, and outdoor access have unignorable advantages for disease control (Shields and Duncan 2009, 10). These concerns about animal welfare and the environmental impact of large-scale production caused a shift towards cage-free and free-range egg production in the 1970s and a push for regulations. More recently, the trend towards organic and local foods has also influenced the egg industry, with many small-scale producers offering eggs produced from heritage breeds and chickens fed an organic diet or are allowed free roaming.

In addition to its economic and environmental impact, the egg industry has also had cultural significance. Between the inexpensive cost and nutrient dense protein

provided by eggs, eggs are an ubiquitous part of American cuisine. Even at brunch joints in Las Vegas with its abundance of alternative gourmet foods, “a good egg, honestly cooked, is what brings in the most customers, and they eat them in staggering quantities” (Burkhard 2005, 2). Eggs are an essential part of American foods and are featured in many classic dishes such as scrambled eggs, cakes, and sauces. The versatility of eggs has allowed them to be used in a wide range of culinary creations, from breakfast to dinner and from savory to sweet. Eggs have become a staple in American diets and have been an integral part of American culture for centuries. The association between Easter and eggs demonstrates that the cultural significance to eggs extends beyond being a staple American food. Eggs are traditionally decorated and used in egg hunts. Additionally, the famous annual White House Easter Egg Roll is a time-honored tradition, which demonstrates the cultural importance of eggs in American society.

The history of egg production in America is a complex and elaborate story that has been shaped by economic, social, and cultural factors. The purpose of chickens and the American egg industry have undergone significant evolution over time, from small family farms to a massive commercial industry. With the expansion of the industry, scrutiny has arisen regarding its impact on animals. Regardless, eggs remain a staple of American society and their importance in cooking is unlikely to diminish any time soon.

References

Bilger, Burkhard. 2005. “The Egg Men.” *In The Kitchen* 81, no. 26 (September): 110–119. https://www.newyorker.com/magazine/2005/09/05/the-egg-men.

Braunschweig-Norris, Jessica. 2005. “The U.S. Egg Industry – Not All It’s Cracked Up to Be for the Welfare of the Laying Hen: A Comparative Look at United States and European Union Welfare Laws.” *Drake Journal of Agricultural Law* 10, no. 3 (2005): 511–39.

Davis, Janet M. 2013. “Cockfight Nationalism: Blood Sport and the Moral Politics of American Empire and Nation Building.” *American Quarterly* 65, no. 3 (September): 549–74. https://doi.org/10.1353/aq.2013.0035.

Rasmussen, Wayne D. 1983. “The New Deal Farm Programs: What They Were and Why They Survived.” *American Journal of Agricultural Economics* 65, no. 5 (1983): 1158–62. https://doi.org/10.2307/1240440.

Shields, Sara, and Ian J.H. Duncan. 2009. “An HSUS Report: A Comparison of the Welfare of Hens in Battery Cages and Alternative Systems.”(2009): 1–28. https://www.wellbeingintlstudiesrepository.org/hsus_reps_impacts_on_animals/18

Zaheer, Khalid. 2015. “An Updated Review on Chicken Eggs: Production, Consumption, Management Aspects and Nutritional Benefits to Human Health.” *Food and Nutrition Sciences* 06, no. 13 (October): 1208–20. https://doi.org/10.4236/fns.2015.613127.

Meet the Author:

Satchel Walton is a sophomore at UNC from Louisville, Kentucky. He is pursuing a major in Anthropology with a minor in Philosophy, Politics and Economics. He took HIST 880, Readings on the Global History of Capitalism with Dr. Benjamin Waterhouse in the fall of 2023, where he wrote a more extended version of this paper (he would be happy to send you that longer version on request, but cut it substantially because he figured few people would want to read the whole thing). He is interested in all sorts of things including the very long scope of history, the evolution of human culture, and how we can lead meaningful lives in the 21st century. You can contact him at satchel@unc.edu.

The Dark Matter of History: What Economic Historians Should Learn from Global Psychological Diversity

Satchel Walton

Class of 2026 Interdisciplinary Studies Major

Research paper written in HIST 880: Readings on the Global History of Capitalism

Mentored by Dr. Benjamin Waterhouse

Abstract: Over the last twenty years, anthropologists and psychologists have been unearthing the fundamental cross-cultural variation in patterns of thought. Researchers such as Joseph Henrich, Michael Muthukrishna, and others in a field they call "historical psychology" have been using vast datasets to examine ways that peoples' psychologies change over time and how that can be a driver of historical change. Using this expanding literature and others working with large historical datasets, I revisit questions that have long bedeviled economic historians like the factors underlying the economic "great divergence" between China and Europe. In this paper, I will propose two new ways this expanding literature could inform the research of economic historians. I do this by providing a deeper explanation for the relationship between wars and subsequent trends in economic inequality, and illuminating how selective migration can lead to population-level differences in preferences for inequality aversion.

Key words: Economic history, historical psychology, cultural evolution,

Psychological Variation

Scholars across fields have, at times, assumed that, on the most basic psychological levels, all people are relatively the same. Of course there is deviation between individuals, but it was assumed that when comparing large groups, the *averages* would be the same. As recently as 2010, Stanford professor Ian Morris published a book we will revisit that takes a stab at the divergence question, and one of his key lines that he repeated in order to support his (mostly) geographically-based thesis was that "people—in large groups—are all much the same." Because chaps (people) are on average the same, according to Morris, what makes history must be due to maps.

Unfortunately for such historical narratives, cross-cultural psychologists and anthropologists in the 21st century, building on earlier work by the likes of Harry Triandis and Geert Hofstede, have come out with a bevy of findings demonstrating deep psychological diversity between populations (Dubner, 2021). This has shaken the universality of the theories found in psychology textbooks, as a 2008 paper found that only 4% of participants in psychological studies were from Asia, Africa, the Middle East or Latin America. These regions have over 85% of the global population but psychological findings made on WEIRD (Western, educated, industrialized, rich, democratic) subjects very frequently fail to replicate in them (Arnett, 2008).

To reference a few examples of these findings, neuroscientists used to think facial processing happened largely in the right hemisphere, but it turns out that is only true in literate people, who are also significantly worse at facial recognition than illiterate people (Huettig and Mishra, 2014). Many visual illusions made in the West turn out not to work on people who grew up in parts of rural Africa, and even what psychologists call the "fundamental attribution error" turns out not to be so fundamental to human psychology at all—it doesn't even exist in East Asians or Russians (who are far from the most

culturally-psychologically distant people from Americans and Europeans) (Henrich, Heinem, and Norenzayan, 2010).

Historians should care because a significant body of research is finding that Morris's claim, that groups of people are broadly the same, is wrong not just on trivial matters like visual illusions, but also on extremely important ones for economic history, like how willing people are to take risks or to collaborate, how generous they are towards strangers versus family members, or how they balance rewards in the present versus future. This is easy to document cross-culturally today: economic games in which people are asked to distribute money under one of many different scenarios show that people are more generous towards anonymous partners and more willing to collaborate with them if they are better integrated in markets, profess belief in all-knowing and punishing deities, or have been exposed to war-related violence. Researchers in the sub-field they call Historical Psychology are beginning to use some novel sources and advanced statistical methods to study how these dispositions, fundamental to how humans respond to economic incentives, change over time (Muthukrishna, Henrich, and Slingerland, 2021).

The most substantial accumulation of evidence so far from this line of investigation is found in Joseph Henrich's *The WEIRDest People in the World* (2020). He musters a broad range

of statistically-based studies to claim that the average Western European gradually psychologically changed from the fall of Rome through the Enlightenment-era to become more individualistic, more fair towards and trusting of strangers, harder working, patient, and positive-sum. This made cooperative institutions like markets, guilds, scientific societies, and eventually democratic governments with the rule of law more likely to emerge. Since the end results of Henrich's analysis include the rise of capitalism, it could have major

implications for economic historians. Economic historians and other academics would do well to analyze his evidence, which includes everything from the long-term fall in interest rates and murder rates over this period as metrics of increasing patience, to statistical correlations between present-day psychological, economic, and demographic metrics and his proposed causal factors like a region's Catholic Church exposure, density of monasteries of particular types, and number of chartered cities. Subsequent studies and reviews have continued to build support for the thesis. For instance, textual analysis of the language in English and French plays from 1600 through 1800 shows a long-term rise in prosocial attitudes as words related to cooperation and trust became more common relative to words related to strength and anger, with especially swift changes prior to the English Civil War and the French Revolution (de Jesus Dias Martins and Baumard, 2020).

Psychological Change and the European Divergence

Henrich's theory to explain these long-term psychological changes is a surprising one: that norms regarding marriage and the family adopted by the Catholic Church at the Council of Elvira in 305 AD, and enforced on Western Europe over the subsequent centuries, broke down the extended family ties and clan networks that had previously dominated societies throughout world history, creating psychologically peculiar people in the process. He uses a variety of data sources to demonstrate that as the Church enforced increasingly strict bans on polygamy, arranged marriages, and especially cousin marriages, clan ties dissolved and people were put in an

environments where they had to rely on their individual traits rather than their family roles. People exposed to the Church's new norms then joined voluntary organizations like guilds, charter cities, and monasteries that gained their loyalties and re-doubled their individualism. By the High

Middle Ages, they became more inclined towards the positive-sum thinking, morality based on internal mental states, impartial rules, and impersonal prosociality that were necessary to eventually create well-functioning markets and representative, law-based governments.

Other researchers have noted the distinguishing psychological characteristics of WEIRD populations and attempted to explain them and their relation to economic development. Most notably, Nicolas Baumard (2019) proposes, based on the biological principles of Life History Theory[2] , that 18th century British living standards became so high that rich Britons began to have different preferences on basic issues: they had become rich enough to indulge in risky projects with uncertain and long-term benefits, and their preferences adjusted to be more optimistic, more socially trusting and cooperative, and more willing to invest in education. This psychological change in turn increased the rate of innovation. Baumard does recognize something that Henrich's model cannot easily explain: cross-Channel social and economic differences that were by the 18th century so favorable to the English that one might accuse Baumard of jingoism for trumpeting them were he not French. The most glaring example he mentions is that life expectancy on the eve of the Industrial Revolution was 24.8 in France and 42.1 in England.

One other well-documented cross-Channel mini-divergence in demographics, however, seems to undermine Baumard's theory.[3] Starting in the early 18th century, fertility went through an unprecedented plummet in poor France while it increased in rich England. The phenomenon of French fertility decline is

[2] LHT seeks to explain the variation in different organisms' timing of life events based on the evolutionary incentives they face. It predicts that organisms in more abundant environments will invest more in long-term, future-oriented development.

[3] Henrich himself gives a rebuttal to Baumard in the endnotes of his book, but it seems half-hearted and inadequate to me. He cites a study showing that the age at menarche is not actually lower in females exposed to domestic abuse when you account for genetic confounding (a marginal part of Baumard's argument anyway), when it seems to me the much more obvious rebuttal would be that the age at

striking and poses a major problem for many conventional theories of the demographic transition based on economic development. It is a fairly basic component of Life History Theory that organisms in richer, more stable environments will have fewer children (and higher investment per child), and one that Baumard himself applies in the paper in demonstrating that rich Britons saw their fertility decline in the 18th century relative to poor Britons. And yet according to a recent study (Blanc, 2023) that uses some of the big data methods of historical psychology, the best explanation for French fertility starting to decline under Ancien Regime semi-feudalism, when life expectancy was still around 24, seems to be historical and religious in nature. Areas that had high religious conflict and were subsequently subject to extractive institutions run by the Counter-Reformation Catholic hierarchy started showing large declines in religiosity in the early 18th century. They had fewer last wills and testaments using religious language and fewer people requesting to be enrolled in the Perpetual Mass held by Catholics for the dead. Then, those same areas showed massive declines in fertility. Findings like this have prompted the demographer Lyman Stone (2023) to say that "all fertility decline everywhere in all places was caused by French people inventing atheism sometime between 1700 and 1735" (he clarified that it was a joke, but only 65% a joke). This is another good example of exogenous religious and cultural factors affecting economic history in important ways: for one, France's fertility decline allowed it to keep near-parity with Britain in per-capita GDP despite not having the Industrial Revolution, and also on a more basic level the decline affected European demographics so much that Braudel wondered if the French lost their European supremacy not at Waterloo, but at some unexplained moment "during the reign of Louis XV when the natural birth-rate was interrupted" (Blanc,

menarche has been secularly *declining* in all rich countries over the last few centuries (Sørensen et al., 2012) when a LHT perspective would predict a later age at menarche for people with higher standards of living.

2023). To return to the point, this is an example that is difficult to square with Baumard's theory that lifestyle change was driven by the changing attitudes brought about by richness in any simple way.

Precedents in the Historical Literature for the Thesis of Long-term Psychological Change

Henrich's thesis about Catholic marriage norms being the source of long-term psychological changes culminating in the Western divergence is certainly not part of the conventional narrative of European history, but other parts of Henrich's narrative about long-term cultural change is adding statistical power from the age of big data to perspectives already well-established within the existing historical debates about European history, capitalism, and the great divergence.

Historians have speculated about deep psychological changes in how people perceive the world throughout history. In Thompson's (1967) account, European perceptions of time changed dramatically sometime around the 18th and early 19th centuries as the rigidities of industrial capitalism, the spread of clocks, and the powerful Protestant conception of thrift led to a shift away from task-orientation to an understanding in which every moment could contain work, could be lost, and should be saved. Despite its wide influence in historical scholarship on the issue, Vanessa Ogle (2019) points out that Thompson uses an "often impressionistic methodology, relying on few, mostly literary sources." Could there be a better data source, though, to try to understand such a monumental and barely perceptible shift like how hard people were working in a society?

Henrich gathers a couple of studies to empirically measure what the academic conversation around Thompson and Ogle had only vague impressions of. For one, data from

London criminal courts show that people testifying were increasingly likely to say

they had been working at the time of a crime as the 18th century progressed, suggesting that total working time increased by 40% from 1748 to 1803. Another study (Clark, 1987) shows that in England, threshing got twice as efficient from the 14th to early 19th centuries, even though the method (whacking grain stalks with a stick until the seeds fell off) did not change at all over that time period, which leads Henrich to conclude that "people were simply working more intensely."

That idea lines up neatly with Jan de Vries's concept of the "Industrious Revolution" (2008) in which Europeans began to willingly choose to work harder in the 17th and 18th centuries. More generally, the picture of European history since the Black Death as being driven by a complex web of factors including the long-term cultural shift from the medieval universities through the Scientific Revolution, Enlightenment, and Industrial Revolution is not new[footnote]. Some recent historical literature, though, has pointed towards a narrative that looks quite similar to Henrich and Baumard's theories on preference changes. For instance, data on the effects of marketplace integration on prosocial behaviors fits right in with the thesis set forth by Deidre McCloskey's *The Bourgeois Virtues* (2006), which asserts, contrary to narratives about markets making people unscrupulous, that the spread of commerce and enrichment of society made people more benevolent. And McCloskey has in mind not just virtues like thrift or hard work, but her four classical and three Christian virtues. One may reasonably doubt that something as intangible as

Extra footnote inserted during editing, sorry format weird: I just flipped through McKay, Hill, and Buckler, 1983 as a representative of textbook Western history from four decades ago to test that theory, and you can obviously see many of the same themes

love or courage has truly increased over time due to market expansion. McCloskey's evidence is often as vague and inconclusive as Thompson's evidence on time was, but studies on

the effect of marketplace integration on impersonal prosociality line up with the

thought that markets do make us nicer: when deciding how to split money with an anonymous stranger, highly market-integrated people (in cultures as diverse as Accra, Ghana and rural Missouri) give close to half, while people with no market integration only give a quarter. The only other factor that was a significant predictor of giving was adherence to a religion with a moralizing and punishing deity (Islam and Christianity, primarily). Mielants (2007) identifies the origins of the European divergence in developments in the Middle Ages, especially chartered cities which had to compete with each other and which Mielants sees as fundamentally different in structure from those in China, India, or the Islamic world.

Some of Landes's arguments in *The Unbound Prometheus* (1969) seem to hold some water in this context. Although he did not understand the underlying mechanisms of psychological change and competition between cities, Landes asserts that living standards grew near-continuously between 1000 and 1750, and sees that a self-reinforcing process had taken hold by the Middle Ages, in which "those economies that were freest seem to have been most creative; creativity promoted growth; and growth provided opportunities for further innovation."

However, just because a divergence narrative with a big role for particular European cultural and then institutional practices has some new data-based firepower does not mean that we must give credit to anyone who previously made a non-geographic divergence argument. One popular book in this category is Niall Ferguson's controversial (and provocatively-titled) 2011 work *Civilization: The West and the Rest*, which argues that the West became prosperous because it developed six "killer apps". If you squint, you can see that Ferguson recognizes some of the correct cloud of

factors to consider (along with others, like modern medicine, that seem clearly

11

downstream rather than upstream of growth): innovation rates, patents, Protestantism, science, and institutions that promote the rule of law all surely played *some* role in European growth over the last five centuries. But the web of causation Ferguson draws is garbled and his history is sometimes flat-out wrong. To take the Reformation as an example, Ferguson writes that "until the Reformation, Christian religious devotion had been seen as something distinct from the material affairs of the world. Lending money at interest was a sin [etc.] … All that had changed after the 1520s" even though more careful studies of the Reformation-era have found it "increasingly difficult to accept the radical post-Reformation discontinuity in economic ethics" first postulated by Weber and Tawney (Waddell, 2012). Ferguson then proceeds to attribute far too much importance to Protestantism in East Asian growth,[footnote] implying that America is immune to European secularization (a narrative that looks sillier with every year of data collected), and then transition to strange speculation on whether Western civilization was about to have its Fall of Rome. Ferguson can string together trite historical anecdotes

Editing footnote 2: It's fine if a sixth of South Koreans are now Protestant, as Landes states, but has he glanced at the figures in, say, Liberia?

in order to rehearse the tropes of some of the divergence literature, but we do not have to give him credit for rehashing the vaguely right constellation of factors and then badly screwing up the web of causality. That's not hard, or interesting.

Landes, despite the parts of his *Unbound Prometheus* discussed above that the data seem to bear out and despite being a much more careful historian than Ferguson, also makes some arguments that do not measure up. Again considering the Reformation, in *The Wealth and Poverty of Nations* (1998)

Landes attributes the success of Protestant economies to their rediscovery, after translating the Bible for themselves, of the strong tradition of property rights supposedly demonstrated by the Israelites in the Old Testament. This seems patently incorrect for

12

a lot of reasons, not least that to whatever extent that Protestants (or Calvinists particularly) got values supporting accumulation of private wealth and property from a book with more direct commands like “Jesus said … go and sell that thou hast, and give to the poor” (Matthew 19:21, KJV) and “Ye cannot serve God and mammon,” (Matthew 6:24, KJV), it must say more about their underlying psychological predispositions than it does about the text.[Footnote]

Henrich’s narrative takes all of the factors Ferguson and Landes identify, and adds a profound depth to them. In contrast to the accounts given by Ferguson and Landes, Henrich’s well-crafted narrative depicts Protestantism as a “booster shot” to the psychological changes already taking place across Western Europe, saying it succeeded because “their core religious values and ways of seeing the world meshed with the era’s proto-WEIRD psychology”—salvation was achieved through *internal mental states*, church governance became

New footnote: For a review of how early Protestant reformers justified the beliefs that led to the Protestant work ethic, see Eaton (2013).

more participatory, callings were to be followed with diligence. Unlike Ferguson or Weber, whose narratives seem chauvinist for their idea of Protestant culture despite the authors’ personal lack of religion, Henrich shows it to be an important, though not unique or entirely positive, influence on subsequent economic and political developments. He explains that areas with Catholic orders founded around the same time—Cisternians, Brethren of the Common Life, Jesuits, which in many ways seemed to adopt the same set of proto-

WEIRD beliefs—look similar to Protestant areas on a variety of metrics, showing trends common across confessional divides, while also showing differences such as how suicide rates became much higher in areas that went Protestant than those that stayed Catholic.[Footnote]

Footnote: And while his book's subtitle makes it clear that it is about "the West," Henrich is careful to note that there is no "the rest" that can be lumped together in any cogent fashion, and he has a section on the nascent research about psychological diversity within India and China which would be longer were most of the existing research not about Europe.

13

At this point, it should be clear that the validity of all sorts of claims made by Henrich, Baumard, Landes, and others rest at least partially on claims of exactly how high English and broader European standards of living in the 18th century were as compared to other world regions and historical eras. Pomeranz claims in *The Great Divergence* (2000) that as late as the mid-18th century, China and Europe (and England and the Yangtze Delta) were still equivalents on important economic metrics—however, he revised his claims in response to subsequent research to set an earlier date on the first notable divergence, acknowledging that "we do need to look at the great divergence as emerging over time, rather than as simply a sudden break" (Pomeranz, 2011). Ian Morris, who has recently advocated more ancient historians take evolutionary perspectives (2022), published in 2010 estimates of long-term social development based on complex metrics of urbanization, military capacity, and energy capture. He found that 18th century Western Europe *was* the most developed region of the world at the time. But, Morris says it was still on the same level of socio-economic development that Song Dynasty China and the Roman Empire had been in their heydays, and theorizes that Europe was able to break through an invisible ceiling that had previously constrained societies that reached that level of development. The heroic national accounting work of Stephen

Broadberry and his colleagues has given us an account of this issue that is, if not definitive, at least the best we have gotten so far.[4] They find that British GDP per capita was the highest the world had ever seen immediately prior to the Industrial Revolution, and that "the transition to modern economic growth built on the earlier foundations of a persistent upward trend in GDP per capita which doubled from 1270 to 1700" (Broadberry et al., 2015). Adding in better Chinese data, Broadberry, Guan, and Li (2018) ultimately come to the boring (but likely intellectually accurate) conclusion that "the Great Divergence began earlier than originally suggested by the California School [i.e. Pomeranz], but later than implied by older Eurocentric writers [i.e. Landes]." Pomeranz's own revisions along with work by Morris and Broadberry all point in more or less the same direction: 18th century Western Europe was the most prosperous region of the world, and likely the most prosperous region ever up to that point, but was not yet radically different from historic precedents such as the Northern Song Period in the 11th century.[5] These results make the idea that some underlying cultural factor like WEIRDness caused Western European growth through the Middle Ages seem possible, and it might be that some unique cultural-evolutionary package allowed Europe to surpass the social complexity of the Song and Romans. But that package, if it existed, was not necessary for societies to achieve levels of prosperity like those Europe had up to the late 17th century.

[4] The Brazilian economist Rafael R. Guthmann (2022, 2023a, 2023b) makes arguments much to the contrary of nearly all these accounts on his blog. He claims that living standards in the Classical period of Ancient Greece were higher than those in Britain on the precipice of the Industrial Revolution based on housing and caloric data. He also claims that the Malthusian trap never really existed, as plenty of societies saw fertility decline once they reached a certain level of affluence. He cites some solid sources so I'd like to look into the evidentiary basis of his claims more, but he's also a random professor in Chile posting on his blog, so he gets relegated to a footnote unless and until he gets this stuff in a peer-reviewed journal.

Because of fundamental disagreements

like "which of these regions was richer?" and the replication crisis in social science fields, any new ideas coming out of anthropology, archaeology, and psychology need scrutiny before being written into the "big history" cannon, and are often inconsistent. Why is it that reading Turchin (2015) and then Henrich (2015 and 2020) gives one the conclusions that war between tribes is pretty much the default prehistoric state of humanity and that the Enlightenment was endogenous to long-term European psychological shifts, whereas reading Graeber and Wengrow (2021) would suggest a generally more peaceful pre-history and an Enlightenment that was essentially just a rip-off of Native American ideas? Clearly there is still much work to be done. But, in the next section, I tentatively outline two examples of ways I think historians of capitalism could benefit from studying the empirically-grounded research on psychological and cultural variation through time. Similarly to Henrich's book, these findings about psychological and cultural change would not overthrow historians' existing narratives, but may be able to elucidate deeper causes for phenomena historians have only partially explained.

[5] Occasionally I am skeptical what the one number of GDP per capita or social development indices can tell us—anyone who had read a world history textbook would be able to tell you that 18th century Europeans never before seen global trade networks, unprecedented levels of exploitation of the Americas and slave plantations, and government institutions in England and Holland that, while certainly not democratic, clearly had the roots of democracy.

Future Directions for Economic Historians

a) What Creates the Economic Effects of Wars?

A crucial part of the narrative of Thomas Piketty's *Capital in the Twenty-First Century* is that economic inequality in Europe (and to a lesser extent the U.S.) declined precipitously because of the historically unprecedented economic and political shocks of the two world wars, and that inequality showed no signs of declining without them. In trying to identify the causes of this fall of inequality in Europe during and in the aftermaths of each of the

world wars, Piketty attributes part of it to the physical destruction of capital during the war, part of it to the low rate of savings and private investment in the period, and part of it to "deliberate policy choice[s] aimed at reducing—more or less efficaciously—the market value of assets and the economic

16

power of their owners" (Piketty, 2013).[6] Why were those deliberate policy choices made across Europe in the aftermath of the world wars?

First, it is worth noting that Piketty claims that the same general course of inequality in the twentieth century—high in the Belle Epoque, falling through the wars, and rising again since 1980—is "a phenomenon that affected all European countries" with a generally similar mix of causes. According to Scheve and Stasavage (2010), top tax rates did *not* increase after World War I in countries which did not mobilize large shares of their population in the war, such as Sweden, the Netherlands, Spain, and Japan. Conversely, Scheve and Stasavage found that top tax rates spiked in countries like France, Britain, Canada, and the United States that mobilized a large share of their populations. The authors also found that neither the expansion of suffrage to all male citizens nor the presence of leftist parties in a nation's parliament led to any appreciable increase in high-income tax rates in the countries that did not experience large-scale mobilizations in World War I. It seems to be a society's experience of war itself that creates a demand for egalitarian policies. This would also fit with Piketty's narrative to some extent, as he notes that changes in U.S. inequality after the wars were much less dramatic than those seen in Europe, and the U.S. had less war exposure. Why, though, would exposure to war lead to greater and more effective demands for redistribution and, consequently, higher top tax rates?

[6] The reliability of the Piketty-Saez data has recently been challenged due to multiple methodological disputes mainly regarding the United States. Auten and Splinter (2019), primarily by adjusting for changing household composition, show top 1% after-tax income shares in the U.S. to be essentially flat since 1960, for example (see also Smith et al. 2019).

The challenge to the Piketty-Saez data from Geloso et al. (2018) could be problematic for the narrative of war experience leading to demands for lower inequality, as their revisions show the precipitous decline in top-income share to be during the Great Depression, not during or following the Second World War. However, the data Piketty cites on the decline of inequality in Europe due to the world wars has gone relatively unchallenged and Europe had much higher per capita mobilization for a longer period of time in both wars as compared to the U.S. Also, Piketty and collaborators continue to fire off in-the-weeds methodological retorts along with spicy public statements defending the validity of their U.S. estimates from these critiques (Saez and Zucman 2020) most recently on Wednesday (Piketty, Saez, and Zucman, 2023). For these reasons, I will take his narrative on the decline of inequality in Britain and France during and after the world wars as probably fairly accurate for the purpose of this paper.

17

It is no secret that the wars catalyzed support for expanded unemployment insurance, health coverage, and social policy broadly, and expanded the moral bargaining power of the people fighting for them. It is no secret that, using Britain as an example, David Lloyd George asked two weeks after Armistice Day "What is our task? To make Britain a fit country for heroes," or that the Beveridge Report gained massive public popularity in the midst of the Second World War. Plenty of historians have recognized that there was a strong popular demand for a more egalitarian society after the wars, made possible and patently necessary by, among other factors, the sacrifices of the common people for the war effort. Tony Judt (2006) wrote that "It was the war that changed all this. Just as World War One had precipitated legislation and social provisions in its wake … so the Second World War transformed both the role of the modern state and the expectations placed upon it." Ta Nehisi Coates (2013) memorably answered the question of why the U.S. couldn't have a welfare system like Western Europe by writing that it was possible by "fighting a genocidal war which results in massive relocations, more ethnic homogeneity, the near-extermination of one of our minorities (one guess at who that would be), and the reduction of our major cities to rubble."[9] But what these discussions lack is a recognition of the partially instinctual and psychological mechanisms underlying changing public opinion on welfare states.

A rapidly expanding body of research from war-torn areas is exploring how people exposed to wars become *psychologically* distinct from others, which

may lead to greater individual preferences for redistribution. In Sierra Leone, for example, researchers ran various "allocation games" with people eight years after the end of that country's civil war. In a basic version of one of these games, a participant would be given a certain amount of money and decide how to split it between themselves and an anonymous partner (who would not know the participant's identity), who the participant was told was either from their own village or from a distant village. When told that the anonymous person was from a distant village, the people exposed to violence in the war and those who were not exposed to violence allocated money in the same way. But when told the participant was from the same village, participants exposed to violence were less selfish and more averse to inequality in the various games than were the people who were not exposed to violence (Bauer et al. 2013). This is part of a broader academic literature that is finding that people exposed to war violence and mass mobilization engage in more prosocial behaviors for in-groups, reviewed by Bauer et al. (2016).[7] With deep roots in human evolutionary psychology (if the in-group of the tribe is experiencing a lot of conflict, it is crucial for it to be socially cohesive), people who experience wars are not just more egalitarian for the in-group, but more likely to adhere to social norms of all sorts, from voting to attending religious services to attending community meetings to ostracizing minorities. Results like these could also provide a partial explanation, too, for social scientific findings such as those in *Bowling Alone* (Putnam, 2000), which show how so many markers of civic engagement and community cohesion rose for a few decades after the Second World War and subsequently

[7] Charlotte Fiedler's more recent literature review calls into question some of Bauer et. al's (2016) more sweeping conclusions, but substantiates that the "increase in cooperation [following exposure to war-related violence] can often be explained by prosocial behavior toward the in-group but not the out-group" (Fiedler 2023).

declined.[8]

These results together suggest that increasing in-group loyalty was not unique to the world wars, and that increased social solidarity and egalitarianism following war is not purely the result of rational actors seeing the sacrifices of soldiers or of a changing political balance of powers between workers and rulers. It is part of a broader and deeply human response in which conflict shifts people's psychologies, leading them to invest more in social connection and increase their in-group solidarity.

[8] The findings also match with the idea that social cohesion would peak a few decades *after* exposure to a conflict, as short-term positive feedback loops keep some prosocial behaviors rising for a period of time after exposure to violence: e.i., your friends feel more inclined to join a church or a bridge club or whathaveyou, which leads the norm to strengthen enough that *you* feel inclined to join, thus strengthening the norm further until some of your neighbors join, although the positive feedback loop cannot go on forever. I'm sure someone has made and discussed more thoroughly the same connection between these findings and *Bowling Alone* before and I should cite them, but I couldn't find anything that does so explicitly.

[9] The increased ethnic homogeneity after each of the world wars—the result of borders being redrawn and people displaced in Eastern Europe and the Holocaust across the continent—probably contributed to the effect increasing in-group prosociality had on support for welfare states. After all, if you had previously seen Jews or other ethnic minorities in your state as out-groups, the out-groups were largely gone by 1945.

Concluding that people's preferences for taxation and redistribution changed because exposure to conflict changed their psychologies by increasing in-group solidarity would not, then, be a radical revision of the traditional historical narrative on the creation of the European welfare state—certainly not as radical as the conclusion that the roots of the European takeoff was ultimately due to early Church leaders' opinions on cousin marriage. Still, it could provide an additional layer of depth to the narratives of historians like Piketty and Judt. A rapidly developing literature on the long-run effects of culturally-selected migration could lead to more novel and unexpected conclusions on the roots of preferences for redistribution.

b) Culturally-Selected Migration, Individualism, and Redistribution

Historians have written extensively about the long-term effects of immigration on a country's culture. In *Albion's Seed: Four*

British Folkways in America, David Hackett Fischer (1989) makes an extensive argument that the deepest aspects of American political culture and regional differences are still based on the distinctive cultural attitudes brought by colonial-era immigration from different types of British people. The tendencies of the Puritans who migrated largely from East Anglia, for instance, are apparently the cultural ancestors who

continue to define New England, parts of the upper Midwest, and American elites in realms like higher education. Views like this have gained some popular credence: Malcolm Gladwell makes this sort of argument in *Outliers* (2008), where he uses psychological studies (Cohen et al., 1996 is a popular one) to draw the conclusion that higher levels of Southern violence can be attributed to the influence of the culture of honor that the Scots-Irish apparently brought to Appalachia.[10] Could it be the case that crucial elements of American culture and politics continue to be influenced by long-dead cultural ancestors from Europe? Some more recent studies of the deep roots of modern divergences do suggest that old folkways survive the American "melting pot" enough to influence the way our distinctly individualistic variety of capitalism works.

Giuliano and Tabellini (2020) find that American counties that received more European immigrants from 1910 to 1930 have (adjusting for covariates) significantly higher levels of support for welfare and redistribution policies today. What's more, support for welfare and redistribution is even higher in those counties that received more immigrants from countries that had larger and more well-established welfare states, and in counties where those immigrants were more thoroughly integrated into society. Overall, they conclude that "immigration left its footprint on American ideology via cultural transmission from immigrants to natives." This is far from confirmation of the narrative Hackett, Fischer, and Gladwell

promote—and it would be much more difficult to do similar studies of an era before there were census records and trying to distinguish between varieties of British immigrants. But it is at least solid empirical confirmation of one of the more controversial ideas: that immigrant preferences can affect the preferences of

[10]I find this particular argument to be prima facie implausible for some reason, and it makes me a bit angry… but then again, the theory does predict that British-descended, Appalachia-adjacent Southern Presbyterians would react angrily to perceived slights against them, so maybe my gut reaction is just confirmation of it.

people other than their descendents at least a century later, and those preferences can affect economic and political realities. Beck Knudsen (2022) uses similarly vast data from censuses and surveys to look at the effects of culturally-selected immigration on the economic preferences on the other side of the Atlantic. Using selection of less common first names and two other proxies for people's level of individualism, she finds that Scandinavian immigrants to North America were significantly more individualistic than those who stayed. Moreover, regions of Scandinavia that sent more of those individualistically-leaning migrants to North America became the most collectivist parts of their countries. Norway, Sweden, and Denmark saw among the highest per capita out-migration of any European countries in the 19th and early 20th centuries; this could help explain why the region is so supportive of high taxation and welfare today: their individualists left, and collectivists stayed.

Similar effects on cultural beliefs and economic preferences can be seen within the United States, with migrants with individualist values being attracted to the frontier and leaving their cultural legacy to this day (Bazzi, Fiszbein, and Gebresilasse, 2020). Adding some nuance to the arguments of historians like Frederick Jackson Turner, the authors of that paper find that Americans who moved to the frontier were more likely to be individualists initially *and* that time spent

on the frontier made them more individualistic still as economic incentives promoted self-reliance and independence. Surprisingly, they found that counties that spent more time on the frontier have citizens today with significantly higher levels of individualism and more opposition to redistributive policies than would be expected based on factors like rurality, race, and political party registration alone (but no difference in opinions on issues unrelated to redistribution, like trade agreements or foreign policy).[11] This historically-determined variable subsequently manifested in increased support for Republican

[11] To get an idea of some of the effect sizes, the time a county spent on the frontier varied between 1 and 63 years, and one extra decade of frontier time was associated with a 3 percentage point higher vote for Trump in 2016, which was roughly the difference in the last election between North Carolina and a

candidates from 2000 through 2016. Olsson and Paik (2016) suggest a more controversial, and empirically more difficult to support, theory based on the same mechanism of migration to frontiers selecting for individualistic as opposed to collectivist people. They note that measures of collectivism such as the importance people assign to obedience generally increase as you go from Scotland and Scandinavia southeast towards the Persian Gulf, and argue that this is because of western agriculture's Middle Eastern origins. Agricultural societies required strong collectivist norms, so as they were established in the Fertile Crescent and gradually expanded outward, more individualist people self-selected into moving further from the core.

The literature on culturally-selected migration is still rapidly developing. Trying to synthesize even the part of it about individualism would present some obvious problems at this point. While Beck Knudsen suggests that immigration to the U.S. made the country more opposed to redistribution because of selection for individualist Europeans, Giuliano and Tabellini suggest the U.S. became more pro-welfare as Europeans brought their pro-redistribution

preferences. Olsson and Paik would place Scandinavia as one of the most individualist regions based on the low importance it places on obedience, while Beck Knudsen sees it as collectivist based on its having some of the world's strongest social safety nets. As with many psychological metrics, there are problems in defining what exactly we are trying to measure and what the proper proxies are for it.

Still, historians of capitalism should pay attention to the above studies. Three of them discuss the potential economic effect of individualism-collectivism, as it is the element

slightly more Republican-leaning state like Iowa or Ohio. For 16 of the 20 presidential elections in the 20th century, frontier exposure had no effect on Republican vote share, but it began to have a strong and growing effect starting in 2000, which the authors attribute to a 21st century divergence between the parties on the frontier-associated values like independence. The effect also manifests in actual policy, with a decade of frontier exposure being associated with as much of a decrease in property tax rates as a 10 percentage point increase in Republican vote share (comparing counties within the same state).

23

of culture that the literature most strongly supports as having a causal effect on long-run economic growth and rates of innovation (Gorodnichenko and Roland, 2012). Economic growth and policies that lead to redistribution are two of the most important phenomena that historians of capitalism must address. These psychological studies may help to address them within a "varieties of capitalism" framework, popular among the new historians of capitalism, that emphasizes how capitalism can exist within and be greatly influenced by a range of different legal, cultural, and social contexts (Rockman, 2014). Psychological or more broadly cultural variation will never be sufficient to explain them alone. Forces like historically contingent movements and events, shifting modes of production, and changing supply and demand for labor will always have a role in any good historical narrative. But if psycho-cultural change can elucidate a fraction of the debate on why different places see different rates of growth and innovation, or why some successfully campaign for mixed market

social welfare capitalism while others get a more laissez faire or corporatist style capitalism, then it is important to pay attention to.

c) Population-level Genetic Differences are Unimportant

It's important to address the concern that studies like this might lead to problematic discussion of *genetic* differences between people. Individualism-collectivism, like really any trait, has some variation due to genetic randomness. If immigrants are, over the course of centuries, selected based on these traits, it raises the question of whether it could have led to genetic differences between populations in traits like individualism-collectivism. Consequently, that raises the question of whether the whole project of cultural evolution could go off the rails and hurtle into territory somewhere between *The Bell Curve* and social Darwinism. A substantial body of research on the 5-HTTLPR S-allele, a polymorphism in a gene relevant to serotonin transporters, purports to have implications for cultural variation (Marcus and Cetin, 2023). Marcus and Cetin claimed to find "a significant role of genetics in predicting cross-societal cultural values

24

variation" based on it, Way and Lieberman (2010) have discussed a potential correlation between it and countries' levels of individualism, and Kong (2015) found that its prevalence in different countries predicts whether democracy creates social trust.

But while there *are* well-documented instances in which cultural/historical differences within the last 10,000 years have led to population-level genetic changes, chronicled by the subfield of culture-gene coevolution (O'Brien and Laland, 2012), most of them are about ancient agricultural developments (Gerbault et al., 2011; Chiao and Blizinsky, 2010).

While it is reasonable to be uneasy about any discussion of group-level genetic and psychological differences given how studies like this are sometimes deployed by racists and cultural supremacists, there are a few reasons not to be exceedingly concerned

about the implications here. For one, our developing understanding of the human genome has, in Henrich's words in an earlier book (2015), "completely dismantle[d] any remaining shreds of the old racial notions," as we can see even more clearly than ever before that racial categorizations have no basis in biology. Second, the most well-established findings of historical phenomena affecting genes are relatively trivial things to do with evolution based on neolithic agricultural history.

25

Henrich's (2015) examples are light-colored eyes (which evolved only in far northern areas that also adopted agriculture) and a mutation that causes alcohol intolerance (which evolved in areas that developed rice-based agriculture). As it stands all serious researchers realize that there is limited evidence for population-level genetic differences directly acting on cultural attitudes, or, for that matter, anything else meaningful to historians.

The LBJ Mistake

In *Lyndon Johnson and the American Dream* (1976), LBJ's former staffer Doris Kearns paints the president as something of a tragic hero whose great downfall is his final years in office: ensconced in the White House, wandering the halls at 3 a.m. agonizing over his decisions, and destroying his Great Society accomplishments by ramping up a disastrous war in Vietnam. In trying to diagnose the flaws in his thinking that led him to drop so many bombs so fruitlessly on North Vietnam, Kearns writes that Johnson often came to shallow conclusions about foreign policy due to his lack of imagination about people outside of America. "Johnson could not envisage a society in which the individual was an aspect of a more comprehensive organism. … [But in Vietnam] individualism was seen as selfish and immoral." He just couldn't imagine that other people, deep down, could have patterns of thought and fundamental desires that were different from those he encountered on Texas ranches or in Senate

hallways. “Although Vietnam was ten thousand miles away, the psychic distance was far greater. So powerful was the American conception of individualism that it resisted even the barest consciousness that another society might conceive of freedom in precisely opposite terms.”

Historians have an even more difficult task than trying to understand people on the other side of the world: trying to understand people who have long since died. Of course, in many ways, people throughout history *are* the same. They are all members of the same species with the

same innate capacities; they have no important differences based on the bogus racial classifications one would pick on the census form; they are all deserving of equal dignity and historians ought to treat them all with respect. Yet these recognitions should not impede historians from taking a hard look at the solid evidence of population-level psychological variation throughout time and space, and the developing literature on the implications of that variation for their fields.

An expanding literature on the incredible breadth of psychological variation throughout time and space has important implications for economic history. I have started to clarify some of them here, such as the effects of selective migration and violent conflict on preferences for economic redistribution, but much more work needs to be done in a variety of social scientific fields. Lyndon Johnson’s blunder came from his inability to understand that his psychological dispositions and cultural values were particular to his own corner of the world, and not universal. Overall, I have argued that, in order to advance their fields, historians must not continue to make the same mistake.

References

Amir, Dorsa, Matthew R. Jordan, Katherine McAuliffe, Claudia R. Valeggia, Lawrence S. Sugiyama, Richard G. Bribiescas, J. Josh Snodgrass, and Yarrow Dunham. "The Developmental Origins of Risk and Time Preferences across Diverse Societies." *Journal of Experimental Psychology: General* 149, no. 4 (2020): 650–61. https://doi.org/10.1037/xge0000675.

Arnett, Jeffrey Jensen. "The Neglected 95%, a Challenge to Psychology's Philosophy of Science." *American Psychologist* 64, no. 6 (2008): 571–74. https://doi.org/10.1037/a0016723.

Auten, Gerald, and David Splinter. "Top Income Shares and the Difficulties of Using Tax Data." *United States Income, Wealth, Consumption, and Inequality*, 2020, 125–52. https://doi.org/10.1093/oso/9780197518199.003.0006.

Bauer, Michal, Alessandra Cassar, Julie Chytilová, and Joseph Henrich. "War's Enduring Effects on the Development of Egalitarian Motivations and in-Group Biases." *Psychological Science* 25, no. 1 (2013): 47–57. https://doi.org/10.1177/0956797613493444.

Bauer, Michal, Christopher Blattman, Julie Chytilová, Joseph Henrich, Edward Miguel, and Tamar Mitts. "Can War Foster Cooperation?" *SSRN Electronic Journal*, 2016. https://doi.org/10.2139/ssrn.2800494.

Baumard, Nicolas. "Psychological Origins of the Industrial Revolution." *Behavioral and Brain Sciences* 42 (2018). https://doi.org/10.1017/s0140525x1800211x.

Bazzi, Samuel, Martin Fiszbein, and Mesay Gebresilasse. "Frontier Culture: The Roots and Persistence of 'Rugged Individualism' in the United States." *Econometrica* 88, no. 6 (2020): 2329–68. https://doi.org/10.3982/ecta16484.

Blanc, Guillaume. "The Cultural Origins of the Demographic Transition in France." *Social Sciences Research Network*, November 22, 2023.

Broadberry, Stephen, Bruce M.S. Campbell, Alexander Klein, Mark Overton, and Bas van Leeuwen. *British economic growth, 1270-1870*. Cambridge: Cambridge University Press, 2015.

Broadberry, Stephen, Hanhui Guan, and David Daokui Li. "China, Europe, and the Great Divergence: A Study in Historical National Accounting, 980–1850." *The Journal of Economic History* 78, no. 4 (2018): 955–1000. https://doi.org/10.1017/s0022050718000529.

Brown, Wendy. *Undoing the demos: Neoliberalism's stealth revolution*. Princeton, NJ: Princeton University Press, 2020.

Chiao, Joan Y., and Katherine D. Blizinsky. "Culture–Gene Coevolution of Individualism–Collectivism and the Serotonin Transporter

Gene." *Proceedings of the Royal Society B: Biological Sciences* 277, no. 1681 (2009): 529–37. https://doi.org/10.1098/rspb.2009.1650.

Clark, Gregory. "Productivity Growth without Technical Change in European Agriculture before 1850." *The Journal of Economic History* 47, no. 2 (1987): 419–32. https://doi.org/10.1017/s0022050700048166.

Coates, Ta-Nehisi. "'War and Welfare Went Hand in Hand.'" *The Atlantic*, November 4, 2013. https://www.theatlantic.com/international/archive/2013/11/war-and-welfare-went-hand-in -hand/281107/.

Cohen, Dov, Richard E. Nisbett, Brian F. Bowdle, and Norbert Schwarz. "Insult, Aggression, and the Southern Culture of Honor: An 'Experimental Ethnography.'" *Journal of Personality and Social Psychology* 70, no. 5 (1996): 945–60. https://doi.org/10.1037/0022-3514.70.5.945.

de Jesus Dias Martins, Mauricio, and Nicolas Baumard. "The Rise of Prosociality in Fiction Preceded Democratic Revolutions in Early Modern Europe." *Proceedings of the National Academy of Sciences* 117, no. 46 (2020): 28684–91. https://doi.org/10.1073/pnas.2009571117.

de Vries, Jan. *The Industrious Revolution: Consumer Behavior and the household economy, 1650 to the present*. Cambridge: Cambridge Univ. Press, 2008.

Dubner, Stephen. "The U.S. Is Just Different — So Let's Stop Pretending We're Not." Episode. *Freakonomics Radio* no. 469, July 14, 2021. https://freakonomics.com/podcast/american-culture-1/.

Ferguson, Niall. *Civilization: The West and the Rest.* New York, NY: Penguin Random House, 2011.

Fiedler, Charlotte. "What Do We Know about How Armed Conflict Affects Social Cohesion? A Review of the Empirical Literature." *International Studies Review* 25, no. 3 (2023). https://doi.org/10.1093/isr/viad030.

Fischer, David Hackett. *Albion's seed: Four British folkways in America*. Oxford University Press, 1989.

Geloso, Vincent, Phillip Magness, John Moore, and Philip Schlosser. "How Pronounced Is the U-Curve? Revisiting Income Inequality in the United States, 1917-1945." *SSRN Electronic Journal*, 2018. https://doi.org/10.2139/ssrn.2985234.

Gerbault, Pascale, Anke Liebert, Yuval Itan, Adam Powell, Mathias Currat, Joachim Burger, Dallas M. Swallow, and Mark G. Thomas. "Evolution of Lactase Persistence: An Example of Human Niche Construction." *Philosophical Transactions of the Royal Society B: Biological Sciences* 366, no. 1566 (2011): 863–77. https://doi.org/10.1098/rstb.2010.0268.

Giuliano, Paola, and Marco Tabellini. “The Seeds of Ideology: Historical Immigration and Political Preferences in the United States.” *NBER Working Papers Series*, 2020. https://doi.org/10.3386/w27238.

Gladwell, Malcolm. *Outliers: The story of success.* New York, NY: Little, Brown and Company, 2008.

Gorodnichenko, Yuriy, and Gérard Roland. “Understanding the Individualism-Collectivism Cleavage and Its Effects: Lessons from Cultural Psychology.” *Institutions and Comparative Economic Development*, 2012, 213–36. https://doi.org/10.1057/9781137034014_12.

Graeber, David, and David Wengrow. *The Dawn of Everything: A New History of Humanity*. New York, NY: Macmillan Publishers, 2021.

Guthmann, Rafael R. “The Malthusian Trap Never Existed.” The Malthusian Trap Never Existed - by Rafael R. Guthmann, February 24, 2022. https://rafaelrguthmann.substack.com/p/the-malthusian-trap-never-existed.

Guthmann, Rafael R. The Industrial Revolution was just catch-up growth, September 9, 2023a. https://rafaelrguthmann.substack.com/p/the-industrial-revolution-was-just?utm_source=p rofile&utm_medium=reader2.

Guthmann, Rafael R. “Fertility and the Life Cycle of Civilizations.” Fertility and the Life Cycle of Civilizations, November 14, 2023b. https://rafaelrguthmann.substack.com/p/fertility-and-the-life-cycle-of-civilizations.

Henrich, Joe, Steven J. Heine, and Ara Norenzayan. “The Weirdest People in the World?” *SSRN Electronic Journal*, 2010. https://doi.org/10.2139/ssrn.1601785.

Henrich, Joseph Patrick. *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton, NJ: Princeton University Press, 2015.

Henrich, Joseph Patrick. *The WEIRDest people in the world: How the west became psychologically peculiar and particularly prosperous*. New York, NY: Farrar, Straus and Giroux, 2021.

Huettig, Falk, and Ramesh K. Mishra. “How Literacy Acquisition Affects the Illiterate Mind – a Critical Examination of Theories and Evidence.” *Language and Linguistics Compass* 8, no. 10 (2014): 401–27. https://doi.org/10.1111/lnc3.12092.

Judt, Tony. *Postwar: A History of Europe Since 1945*. New York, NY: Penguin Press, 2006.

Kearns, Doris. *Lyndon Johnson and the American Dream.* New York, NY: Harper & Row, Publishers, 1976.

Knudsen, Anne Sofie. “Those Who Stayed: Individualism, Self-Selection and Cultural Change during the Age of Mass Migration.” *SSRN Electronic Journal*, 2019. https://doi.org/10.2139/ssrn.3321790.

Kong, Dejun Tony. “A Gene–Environment Interaction Model of Social Trust: The 5-HTTLPR S-Allele Prevalence as a Moderator for the Democracy–Trust Linkage.” *Personality and Individual Differences* 87 (2015): 278–81. https://doi.org/10.1016/j.paid.2015.08.028.

Landes, David S. *The unbound prometheus*. London: Cambridge University Press, 1969.

Landes, David S. *The wealth and poverty of nations*. New York: W. W. Norton & Company, 1998.

Lenin, Vladimir Il’ich. *Imperialism, the highest stage of capitalism; a popular outline*. 2nd ed. London: Martin Lawrence, 1934.

Marcus, Justin, and Ecesu Cetin. “Genetic Predictors of Cultural Values Variation between Societies.” *Scientific Reports* 13, no. 1 (2023). https://doi.org/10.1038/s41598-023-34845-x.

Martins, Mauricio de Jesus Dias, and Nicolas Baumard. “The Rise of Prosociality in Fiction Preceded Democratic Revolutions in Early Modern Europe.” *Proceedings of the National Academy of Sciences* 117, no. 46 (2020): 28684–91. https://doi.org/10.1073/pnas.2009571117.

McCloskey, Deirdre N. *The bourgeois virtues: Ethics for an age of Commerce*. Chicago, IL: University of Chicago Press, 2006.

McKay, John P., Bennett D. Hill, and John Buckler. *A history of western society*. 2nd ed. Vol. 2. Boston, MA: Houghton Mifflin, 1983.

Mielants, Eric. *The origins of capitalism and the “Rise of the west.”* Philadelphia, PA: Temple University Press, 2007.

Morris, Ian. *Why the west rules for now: The patterns of history and what they reveal about the future*. London: Profile Books, 2011.

Morris, Ian. “Evolutionary History.” *Evolutionary Psychology* 20, no. 1 (2022): 147470492110682. https://doi.org/10.1177/14747049211068279.

Muthukrishna, Michael, Joseph Henrich, and Edward Slingerland. “Psychology as a Historical Science.” *Annual Review of Psychology* 72, no. 1 (2021): 717–49. https://doi.org/10.1146/annurev-psych-082820-111436.

Newson, Lesley, and Peter J. Richerson. “Why Do People Become Modern? A Darwinian Explanation.” *Population and Development Review* 35, no. 1 (2009): 117–58. https://doi.org/10.1111/j.1728-4457.2009.00263.x.

O’Brien, Michael J., and Kevin N. Laland. “Genes, Culture, and Agriculture.” *Current Anthropology* 53, no. 4 (2012): 434–70. https://doi.org/10.1086/666585.

Ogle, Vanessa. “Time, Temporality and the History of Capitalism.” *Past & Present* 243, no. 1 (2019): 312–27. https://doi.org/10.1093/pastj/gtz014.

Olsson, Ola, and Christopher Paik. “Long-Run Cultural Divergence: Evidence from the Neolithic Revolution.” *Journal of Development Economics* 122 (2016): 197–213. https://doi.org/10.1016/j.jdeveco.2016.05.003.

Piketty, Thomas. *Capital in the Twenty-First Century*. Cambridge, MA: Harvard University Press, 2013.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman. “Technical Note No. 2023/09: Comment on Auten and Splinter.” *World Inequality Lab*, (2023). December 13, 2023.

34

https://wid.world/document/comment-on-auten-and-splinter-2023-wid-world-technical-n ote-2023-09/

Pomeranz, Kenneth. *The Great Divergence: Europe, China, and the making of the modern world economy*. Princeton, NJ: Princeton University Press, 2000.

Pomeranz, Kenneth. “Ten Years after: Responses and Reconsiderations.” *Historically Speaking* 12, no. 4 (2011): 20–25. https://doi.org/10.1353/hsp.2011.0051.

Putnam, Robert. *Bowling alone: the collapse and revival of American community*. Simon & Schuster, 2000.

Rockman, Seth. “What Makes the History of Capitalism Newsworthy?” *Journal of the Early Republic* 34, no. 3 (2014): 439–66. https://doi.org/10.1353/jer.2014.0043.

Saez, Emmanuel, and Gabriel Zucman. “Trends in US Income and Wealth Inequality: Revising after the Revisionists.” *NBER Working Paper Series*, 2020. https://doi.org/10.3386/w27921.

Scheve, Kenneth, and David Stasavage. “The Conscription of Wealth: Mass Warfare and the Demand for Progressive Taxation.” *International Organization* 64, no. 4 (2010): 529–61. https://doi.org/10.1017/s0020818310000226.

Schulz, Jonathan F. “Kin Networks and Institutional Development.” *The Economic Journal* 132, no. 647 (2022): 2578–2613. https://doi.org/10.1093/ej/ueac027.

35

Smith, Matthew, Owen Zidar, and Eric Zwick. “Top Wealth in America: New Estimates and Implications for Taxing the Rich.” *NBER Working Papers Series*, October 2021. https://doi.org/10.3386/w29374.

Sørensen, Kaspar, Annette Mouritsen, Lise Aksglaede, Casper P. Hagen, Signe Sloth Mogensen, and Anders Juul. “Recent Secular Trends in

Pubertal Timing: Implications for Evaluation and Diagnosis of Precocious Puberty." *Hormone Research in Paediatrics* 77, no. 3 (2012): 137–45. https://doi.org/10.1159/000336325.

Stone, Lyman. "all fertility decline everywhere" *Twitter*, December 6, 2023. https://twitter.com/lymanstoneky/status/1732482928062378032

Thompson, E. P. "Time, Work-Discipline, and Industrial Capitalism." *Past and Present* 38, no. 1 (1967): 56–97. https://doi.org/10.1093/past/38.1.56.

Turchin, Peter. *Ultrasociety: How 10,000 Years of War Made Humans the Greatest Cooperators on Earth*. Champlin, CT: Beresta Books, 2015.

Waddell, Brodie. *God, duty and community in English Economic Life, 1660-1720*. Woodbridge, Suffolk: Boydell Press, 2012.

Way, Baldwin M., and Matthew D. Lieberman. "Is There a Genetic Contribution to Cultural Differences? Collectivism, Individualism and Genetic Markers of Social Sensitivity." *Social Cognitive and Affective Neuroscience* 5, no. 2–3 (2010): 203–11. https://doi.org/10.1093/scan/nsq059.

Meet the Author:

Riley Harper is a third-year undergraduate student at the University of North Carolina at Chapel Hill, where he is pursuing a major in Statistics and Analytics. Riley has a profound interest in the fields of psychology and artificial intelligence, particularly in their application to statistical methods. His academic journey is marked by a recent contribution to the field through his paper on "Distress Tolerance as a Predictor of Early Treatment Dropout in a Residential Substance Abuse Treatment Facility." This work, conducted under the guidance of Dr. Stacey Daughters in the BRANE Lab, revisits and validates the statistical methodologies employed in Dr. Daughters' seminal research, utilizing the original dataset for analysis.

Outside of his academic pursuits, Riley is deeply involved in entrepreneurial and leadership roles on campus. He is the CEO of Chapel Thrill Escapes, an innovative escape room experience designed by students for the UNC-Chapel Hill community. Additionally, Riley serves as the Director of Partnerships and Outreach for AI@UNC, a revitalized club dedicated to harnessing artificial intelligence for academic research enhancement. Under his leadership, AI@UNC focuses on expanding the use and reliability of research sources for faculty members across disciplines.

An Analysis of Distress Tolerance as a Predictor of Early Treatment Dropout in a Residential Substance Abuse Treatment Facility

Riley Harper
riley.harper@unc.edu

Abstract: This paper provides a detailed analysis of an original study conducted by Dr. Daughters on the impact of distress tolerance on early dropout rates from residential substance abuse treatment programs. Employing discrete event history analysis, the paper focuses on the statistical methodologies, including the Cox Proportional Hazards model, to understand the predictive value of psychological and physical distress tolerance measures on treatment outcomes. Our analysis replicated and extended the original study's methods, incorporating rigorous statistical tests to validate the model assumptions and to ensure the robustness of the findings. Through examination, psychological distress measures, particularly the Mirror Tracing Persistency Task (MTPT), were identified as significant predictors of treatment duration, underscoring their importance in predicting treatment dropout. This research contributes to the field by rigorously testing and confirming the original study's findings and by highlighting the utility of Cox Proportional Hazards models in psychological research. The interdisciplinary approach, combining statistical analysis with psychological insights, offers a comprehensive understanding of the factors influencing treatment adherence in substance abuse programs.

Keywords: Distress Tolerance, Substance Abuse Treatment, Cox Proportional Hazards Model, Treatment Dropout, Survival Analysis, Early Dropout Prediction, Statistical Methods in Psychology.

Table of Contents

1. Introduction

In the following pages, I will delve into the intricacies of an original study in the field of substance abuse treatment, offering a detailed analysis and reflection on the methodologies employed. This analysis covers the statistical tests used, their purposes, discusses the relevant R packages and functions, and interprets the outcomes. My primary goal is to showcase the knowledge and skills I've acquired in discrete event history analysis, through the lens of Dr. Daughters' research.

2. Original Study Synopsis

2.1 Research Question

The original study [Daughters et al., 2005] aimed to explore the complex relationship between distress tolerance and early dropout rates in residential substance abuse treatment programs. It focused primarily on assessing how distress tolerance predicts early treatment dropout, and secondarily on understanding the disparity between the influences of psychological and physical distress tolerance on these rates. This investigation was anticipated to yield valuable insights, informing the development of treatment strategies to enhance their effectiveness and ultimately reduce premature treatment dropout.

2.2 Patient Demographics

This study involved 122 individuals enrolled at the Salvation Army Harbor Light residential substance abuse treatment center in Northeast Washington, DC. The average age of participants was 40.3 years, with a majority of 70.5% being male and 95.1% identifying as African American. Educational backgrounds varied: 27.0% had less than a high school diploma, 43.4% had completed high school or equivalent, 20.5% had some college or technical education, and 9.1% held a college degree or higher. Substance use patterns showed 60.7% used crack/cocaine, 41.0% alcohol, 27.9% heroin, and 27.0% cannabis on a weekly basis in the past year. Treatment durations signed for by the participants ranged from 30 days (45.1% of participants) to 180 days (25.4%).

2.3 Recruitment and Screening Process

Out of 144 candidates initially approached within their first week at the center, 16 declined participation. Six more were excluded due to psychosis, determined using the Structured Clinical Interview for DSM–IV (SCID-IV) [Kübler, 2013]. The time from arrival at the facility to study participation averaged 2.6 days. All participants had to demonstrate drug abstinence and complete detoxification before joining the treatment center.

2.4 Study Procedures and Instruments

The study combined the SCID-IV with a series of seven self-report questionnaires and four distress tolerance tasks, with varied order across participants. Performance in the tasks influenced compensation, ranging from $5 to $15. Key tools included the Positive and Negative Affect Schedule (PANAS) [Tran, 2013] for mood assessment, the Center for Epidemiological Studies–Depression Scale (CES-D) [Herge et al., 2013] for depressive symptoms, the Multidimensional Personality Questionnaire–Stress Reaction Subscale (MPQ-SR) [Patrick and Kramer, 2017] for stress-related traits, and the Eysenck Impulsiveness Scale [Huang, 2022] for impulsivity levels. All showed acceptable to high reliability (PANAS: .89, CES-D: .76, MPQ-SR: .84, Eysenck: .76). Additionally, a polydrug use questionnaire assessed drug use across 10 categories. The Interpersonal Support Evaluation List (ISEL) evaluated perceived functional support (reliability: .74), and the Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES) measured readiness to change in substance use behavior (reliability: .69, .85, and .90 for different scales).

2.5 Distress Tolerance Tasks

Distress Tolerance Tasks are pivotal in the empirical investigation of psychological resilience and coping mechanisms. These tasks provide a standardized method to quantifiably measure how individuals withstand and respond to stressful or challenging situations. The empirical data gathered from such tasks are invaluable, as they offer objective insights into human behavior under stress. This, in turn, aids in the development of targeted interventions and therapeutic strategies. Overall, the results we derive from Distress Tolerance Tasks contribute significantly to our understanding of various psychological conditions, particularly those related to anxiety and stress disorders. By employing these tasks, researchers can establish a more robust foundation for psychological theories and practices, ensuring they are grounded in observable, measurable phenomena.

2.5.1 Psychological Stressor 1: Paced Auditory Serial Addition Task (PASAT)

This task [Tombaugh, 2006] involved participants adding numbers that were sequentially displayed on a computer screen, with each new number requiring addition to the one prior. The task was structured in three levels of increasing difficulty, characterized by shorter intervals between number presentations. Participants were informed they could opt out of the final level at any time, with their earnings from the session being contingent on their task performance. Unknown to the participants, the task was designed to end after 7 minutes. Distress tolerance was gauged by the latency in seconds until task termination. Additionally, dysphoria was measured using a four-item scale assessing anxiety, difficulty concentrating, irritability, and frustration, showing acceptable reliability (0.69). To evaluate the increase in psychological stress, dysphoria levels were recorded both at the session's start and after the second PASAT level, thus avoiding confounds with termination latency.

2.5.2 Psychological Stressor 2: Computerized Mirror-Tracing Persistence Task (MTPT-C)

This task [Brown et al., 2018] required participants to trace a red dot along a star using a computer mouse, which was programmed to move the dot in the opposite intended direction. Deviations from the path or a stall of more than 2 seconds triggered a loud buzzer, resetting the dot to the starting position. Participants were allowed to terminate the task at any point, with their performance impacting their financial compensation. The task was secretly set to conclude after 5 minutes. Distress tolerance was again measured by the time taken to terminate the task. Additionally, the number of errors per second was recorded to control for skill level in persistence. Since the MTPT only involved a single level, dysphoria could not be assessed without confounding termination latency.

2.5.3 Physical Stressor 1: Breath Holding (BH)

This task, previously shown to predict cessation attempts in smokers [Brown et al., 2002], was used as a physical challenge. Participants were instructed to hold their breath for as long as possible while an experimenter timed the duration using a stopwatch. Persistence was quantified as the time in seconds before taking a breath.

2.5.4 Physical Stressor 2: Cold Pressor Task (CPT):

This task [von Baeyer et al., 2005] involves immersing the participant's nondominant hand and forearm in ice water (33° Fahrenheit; SD 1°) up to a marked point. Participants were instructed to keep their hand submerged for as long as they could, with the option to remove it at any time. Unbeknownst to them, the task was set to end after 5 minutes. Persistence was measured by the duration before removing the hand from the water.

2.6 Group Status

Participants were grouped based on the duration of their treatment adherence, irrespective of the initially planned length of stay. Early dropouts, comprising 20 individuals, were those who did not complete the initial 30-day treatment period. This group included participants who, against medical advice, chose to voluntarily exit the program, numbering eight, and those dismissed for substance use during the treatment, totaling twelve. In contrast, completers were defined as individuals who successfully met or exceeded the 30-day treatment threshold. This group consisted of 102 participants who adhered to the program's minimum duration requirement, reflecting a commitment to the prescribed treatment course. A further discussion of results will be carried on in the results section of this paper.

3. Methods

3.1 Outline

My analysis aimed to replicate and extend the methodologies used in the original paper to deepen my understanding of the employed techniques. Initially, I focused on recreating the descriptive statistics presented in the study, including the mean and standard deviation calculations for key variables. Following this, I conducted correlation tests to examine linear relationships among the covariates. To further explore these relationships, I performed Welch's Two Sample t-tests to assess the significance of group mean differences. The final step involved implementing a stepwise approach in a Cox Proportional Hazards model, which allowed me to identify the most significant factors among the four psychological and physical tasks discussed in the original paper.

3.2 Descriptive Statistics

First, we will look into the use of descriptive statistics [Anderson et al., 2023], namely the mean and standard deviation. Descriptive statistics are an essential tool in statistics, providing a simple summary about the sample at hand. They do this by simplifying large amounts of data into measures of central tendency and variability. Measures of central tendency describe the center point of a data set–the mean$_{(1)}$, median$_{(2)}$, and mode$_{(3)}$. The mean provides the average value, the median denotes the middle value, and the mode represents the most frequently occurring value. In general, for a set of data, X, with n elements ordered from least to greatest,

$$\textbf{Mean} = \bar{x} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{1}$$

$$\textbf{Median} = \begin{cases} x_{\frac{n+1}{2}}, & \text{if } n \text{ is odd} \\ \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}, & \text{if } n \text{ is even} \end{cases} \tag{2}$$

$$\textbf{Mode} = \underset{x \in \mathbb{R}}{\operatorname{argmax}} \sum_{i=1}^{n} \begin{cases} 1, & \text{if } x = x_i \\ 0, & \text{if } x \neq x_i \end{cases} \tag{3}$$

The other central components of variability describe the spread or dispersion of the data through the range$_{(4)}$, variance$_{(5)}$, and standard deviation$_{(6)}$. The range is the difference between the highest and lowest values. Variance and standard deviation are measures of how far individual data points deviate from the mean. When computing the variance

and standard deviation of a sample rather than a population $n-1$ is used rather than n. This is known as Bessel's correction and corrects for the bias in the estimation of the population variance and standard deviation.

$$\textbf{Range} = \max(x_i) - \min(x_i) \tag{4}$$

$$\textbf{Variance} = s^2 = \begin{cases} \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{j=1}^{n} x_j\right)^2, & \text{if each observation is equally likely} \\ \sum_{i=1}^{n} p_i \cdot \left(x_i - \sum_{j=1}^{n} p_j x_j\right)^2, & \text{if observations are not equally likely} \end{cases} \tag{5}$$

*p_i represents the probability of each observation

$$\textbf{Standard Deviation} = s = \begin{cases} \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{j=1}^{n} x_j\right)^2}, & \text{if each observation is equally likely} \\ \sqrt{\sum_{i=1}^{n} p_i \cdot \left(x_i - \sum_{j=1}^{n} p_j x_j\right)^2}, & \text{if observations are not equally likely} \end{cases} \tag{6}$$

*p_i represents the probability of each observation

The above variance and standard deviation are given for a sample and thus have been denoted with s rather than for the population, in which case they would have been denoted σ. These six measures used in descriptive statistics provide key insights into the central tendency and variability of your data.

3.3 Correlation Test

Understanding the relationship between two variables is often a key objective. One of the most common measures used to assess the strength and direction of a linear relationship between two variables is the sample correlation coefficient$_{(7)}$. This coefficient r_{XY}, typically denoted as r, for two random variables X and Y can be found as follows,

$$\textbf{Correlation Coefficient} = r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{7}$$

Often in software optimizing for computational efficiency requires the avoidance of computing the mean. The following rearrangement of the correlation coefficient$_{(8)}$ allows for this computation to be done without directly calculating the mean,

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}} \tag{8}$$

The Correlation Coefficient computed above is the most commonly used and is known as the Pearson Correlation Coefficient (PCC). There have also been other correlation coefficients developed–such as Spearman's rank correlation–which

are more robust methods and have the potential to demonstrate non linear correlation between variables.

3.4 Student's t-Test

The Student's t-test [The Editors of Encyclopaedia Britannica, 2023] is a fundamental tool for comparing the means of two groups. This test is instrumental in determining whether the observed differences between these groups are significant or the result of random variation. At it's core, the t-test operates under the principle of hypothesis testing. It helps decide whether data can convincingly show a significant difference between two sets of numbers, such as measurements, scores, or observations. The term "Student" refers to the pseudonym used by William Sealy Gosset, who developed the test.

The t-test assumes that the data follows the Student's t-distribution. This assumption is crucial when the scaling term, affecting the data's spread or dispersion, is unknown and must be estimated from the data itself. Several types of t-tests exist, each used in different situations, including the one-sample t-test, independent two-sample t-test, and paired sample t-test. A common application of the t-test is in experimental design, where it helps to determine if a treatment or condition has a statistically different effect from a control condition.

Each of the following tests involve the use of t-scores. A t-score table can be used to mark critical values for different levels of significance, α, and different degrees of freedom, DF. A t-score table is pictured in the appendix.

In the one sample t-test we're often looking to confirm the null hypothesis $H_0 : \mu = 0$ against the alternative hypothesis $H_a : \mu \neq 0$.

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \tag{9}$$

*DF $= n - 1$

Alternatively, when given two groups or variables we may consider the independent two-sample t-test which assumes the two sample sizes are equal and that the two distributions have equal variance. This t-test is defined as,

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{s_{\bar{x}}^2 + s_{\bar{y}}^2}} \tag{10}$$

In the case when there may be unequal sample sizes but the variances of the two distributions are similar such that $\frac{1}{2} < \frac{s_X}{s_Y} < 2$ we can use the pooled variance. When you have two independent samples with similar variances, the pooled variance, denoted as s_p^2, is used as an estimate of the common variance of the two populations. The pooled variance is calculated as follows:

$$s_p^2 = \frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2} \tag{11}$$

*s_X^2, s_Y^2 are the sample variances of the two samples, and n_X, n_Y are the sample sizes

The t-statistic for testing the difference between two means, assuming similar variances but with unequal sample sizes, is given by:

$$t = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \tag{12}$$

*$\bar{x}, \bar{y}$ are the sample means of the two samples, and s_p is the pooled standard deviation–the square root of the pooled variance s_p^2

*DF $= n_X + n_Y - 2$

In the case when there may be unequal sample sizes and unequal variances such that $s_X > 2s_Y$ or $s_Y > 2s_X$ we can use the Welch's t-test [West, 2021]. For this test, we'll utilize $s_{\bar{\Delta}}$ to estimate the standard error of the difference between the two sample means. The formula for $s_{\bar{\Delta}}$ is given by,

$$s_{\bar{\Delta}} = \sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}} \tag{13}$$

Then, the Welch's t-test can be calculated as follows,

$$t = \frac{\bar{x} - \bar{y}}{s_{\bar{\Delta}}} \tag{14}$$

*DF can be calculated using the Welch-Satterthwaite equation

The Welch-Satterwaite equation is used to approximate degrees of freedom without an assumption that the underlying population variances, σ_i^2, are equal and is also known as the pooled degrees of freedom. This equation is given by,

$$\text{DF} \approx \frac{\left(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}\right)^2}{\frac{\left(\frac{s_X^2}{n_X}\right)^2}{n_X - 1} + \frac{\left(\frac{s_Y^2}{n_Y}\right)^2}{n_Y - 1}} \tag{15}$$

3.5 Cox Proportional Hazards Model

The Cox Proportional Hazards model is a cornerstone of survival analysis, a field of statistics that focuses on the expected duration of time until an event of interest occurs. Introduced by Sir David Cox in 1972 [Cox, 1972], this model has since become one of the most widely used methods for analyzing survival data, particularly in the context of medical research and reliability engineering.

Unlike traditional regression models that predict a numeric value or classification, the Cox model is designed to understand and quantify the effect of various covariates on the hazard, or the instantaneous risk of experiencing the event at a given time, conditional on survival until that time. One of its most defining features is the assumption of proportional hazards–the ratios of the hazards for any two individuals are constant over time, regardless of the value of the survival time.

This section aims to delve into the fundamentals of the Cox Proportional Hazards model by discussing how the model is formulated, interpreted, and how it can be applied to real-world datasets to glean insights into factors that influence the time until an event occurs.

Before delving into the Cox Proportional Hazards model we will first define a much simpler model called the Kaplan Meier Model which is the foundational model used in survival analysis. The Kaplan-Meier estimator is defined as,

$$\hat{S}(t) = \prod_{i:t_i \leq t} (1 - \frac{d_i}{n_i}) \tag{16}$$

*t_i are the times when at least one event has occurred

*d_i is the number of events at time t_i

*n_i is the number of individuals known to have survived up to time t_i and are still at risk

$\hat{S}$(t) can be interpreted as the estimated probability that an individual will remain event-free beyond a certain point–t. The Kaplan-Meier curve is thought of as a stepwise function since it decreases at each time point where an event occurs and is assumed to be constant between these points–it is a decreasing function. The estimator is also non-parametric such that it makes no assumption of the underlying distribution for survival times.

With an understanding of the Kaplan-Meier estimator in place, we now turn our attention to the Cox Proportional Hazards Model, a more advanced tool in survival analysis. While the Kaplan-Meier estimator is useful for estimating survival probabilities, it does not allow for the direct assessment of how specific factors or covariates might affect survival times. This is where the Cox Proportional Hazards Model becomes more useful.

The Cox Proportional Hazards model is a semi-parametric model used to investigate the effect of several variables on the survival time of subjects. One of the key features of this model is its ability to handle both categorical and continuous variables, providing a way to understand how different factors contribute to the hazard–the instantaneous risk of experiencing the event of interest. The model is given by the formula,

$$h(t) = h_0(t) \exp\left(\sum_{i=1}^{p} \beta_i x_i\right) = h_0(t) \cdot e^{\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p} \tag{17}$$

*$h_0(t)$ is the baseline hazard which denotes the hazard rate for a baseline level of covariates

*$\beta_1, \beta_2, \cdots \beta_p$ are the coefficients of the covariates

*$x_1, x_2, \cdots x_p$ are the covariates

The baseline hazard quantifies the instantaneous risk of an event occurring at time t, assuming no influence from the covariates such that all covariates are zero or their baseline level. The baseline hazard is demonstrated below,

$$h(t) = h_0(t) \exp\left(\sum_{i=1}^{p} \beta_i(0)\right) = h_0(t) \exp(0) = h_0(t) \tag{18}$$

As previously stated the model is semi-parametic. This is due to the model not specifying a function for the baseline hazard and assuming a multiplicative effect from the covariates on the hazard rate.

3.6 Log Rank Test

The Log Rank Test [Goel et al., 2010] is a non-parametric test used in survival analysis for comparing the survival distributions of two or more groups. The primary application of the Log Rank Test is to assess whether there are statistically significant differences between the survival curves of different groups, often in the context of clinical trials. This test is based on the survival probabilities estimated at various time points, typically when an event (like death, failure, or relapse) occurs.

The key principle behind the Log Rank Test is to compare the observed number of events in each group at each time

point with the expected number of events, assuming no difference between the groups. The test statistic is then derived from the sum of these comparisons over all time points.

The formula for the Log Rank Test statistic is as follows:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_{1i} - E_{1i})^2}{E_{1i}} + \frac{(O_{2i} - E_{2i})^2}{E_{2i}} \tag{19}$$

*O_{ji} is the observed number of events in group j at time i,

*E_{ji} is the expected number of events in group j at time i,

*k is the number of time points.

In practice, the Log Rank Test involves constructing a table that captures the number of subjects at risk, the number of events, and the expected events under the null hypothesis for each time point. Next, the test calculates the total observed and expected events across all time points for each group. The test statistic is compared against a chi-square distribution to determine the p-value, which indicates the probability of observing such a difference if the null hypothesis– which states that there is no difference in survival between groups–were true.

An assumption of the Log Rank Test is that of proportional hazards, implying that the ratio of the hazard functions of the groups being compared is constant over time. However, this assumption does not imply that the hazard functions themselves are constant over time.

The Log Rank Test is favored for its simplicity and the minimal assumptions it makes about the data, and it is particularly powerful in large sample sizes or when the event of interest does not occur often. However, it is less sensitive when the hazard rates cross over time, which may occur in some practical scenarios. In this case, another test which can be used is the Wilcoxon-Breslow-Gehan test. Overall, the Log Rank Test offers a straightforward method to compare survival times across different groups.

3.7 Checking Linearity

Checking linearity is also a critical step in survival analysis, especially when using models like the Cox Proportional Hazards model, which assumes a linear relationship between the log of the hazard function and the covariates. This assumption of linearity influences the interpretation of the model's coefficients along with the validity of its predictions. Verifying this assumption is important to ensure the accuracy and reliability of the model's outcomes.

One common method for checking linearity in the Cox model is through examining Martingale residuals [Gillespie, 2006]. Martingale residuals are the differences between the observed and expected numbers of events, given the covariates for each individual. By plotting these residuals against each of the covariates, one can visually inspect for any systematic patterns that would indicate non-linearity. If the relationship is linear, the plot should show a random scatter of points around zero without any discernible pattern.

Martingale Residuals can be calculated as,

$$\textbf{Martingale Residual} = \delta_i - \hat{H}_i(t) \tag{20}$$

*δ_i is the event indicator (1 if the event occurred, 0 otherwise)

*$\hat{H}_i(t)$ is the cumulative hazard function estimated at the observed survival time for the ith individual

The equation for the estimated cumulative hazard function [Bradburn et al., 2003], $\hat{H}_i(t)$, in the context of the Cox Proportional Hazards model involves integrating the hazard rates over time. This can be expressed as follows:

$$\hat{H}_i(t) = \int_0^t \hat{h}_i(u)\, du \tag{21}$$

*$\hat{h}_i(u)$ represents the estimated hazard function for the ith individual at a given time u

*u ranges from the start of observation at 0 up to time t

*$\int_0^t$ indicates that we are summing up the hazard rates over the interval from 0 to t

This cumulative hazard function is not directly computed as a closed-form integral because the Cox model does not specify the baseline hazard function $h_0(t)$ explicitly. Instead, it is estimated using the data and the proportional hazards assumption.

To further test linearity another approach involves adding interaction terms between covariates and a function of time, such as log(time), to the Cox model. If these interaction terms are statistically significant, this may indicate a violation of the linearity assumption.

In cases where linearity is violated, transforming covariates to instead use the log or square root transformations is a viable approach to achieve linearity. These transformations can be particularly useful with skewed data or when the relationship between the covariates and the log hazard is not linear. After transforming the covariates, it is important to recheck the model to ensure that the transformed variables now exhibit a linear relationship with the log hazard.

Overall, checking linearity in survival analysis involves using residual plots, adding interaction terms, applying transformations to covariates, and reevaluating the model. This process helps validate the assumptions of the Cox Proportional Hazards model and ensures its interpretations and predictions are accurate.

3.8 Checking Proportional Hazards Assumption

The proportional hazards assumption is a key aspect of the Cox Proportional Hazards model. This assumption states that the hazard ratios for different levels of a covariate will be constant over time. Violation of this assumption can lead to biased estimates and incorrect conclusions so several techniques can be employed to evaluate its validity.

A common approach to checking the proportional hazards assumption is to use the Schoenfeld residuals. These residuals represent the difference between observed and expected covariate values at each event time. By plotting the Schoenfeld residuals against time, one can assess if the effect of the covariates changes over time, which would violate the proportional hazards assumption. A pattern which is not random in these plots suggests that the proportional hazards assumption may not hold.

Schoenfeld Residuals for covariate j can be calculated as:

$$\textbf{Schoenfeld Residual for Covariate j} = (x_{ij} - \bar{x}_j(t))(\delta_i - \hat{h}_i(t)) \tag{22}$$

*x_{ij} is the value of the jth covariate for the ith individual

*$\bar{x}_j(t)$ is the average value of the jth covariate for individuals at risk at time t

*δ_i is the event indicator (1 if the event occurred, 0 otherwise)

*$\hat{h}_i(t)$ is the estimated hazard at time t for the ith individual

Additionally, a visual inspection of survival plots for various subgroups within the study can also offer valuable insights on this assumption. This method entails comparing the survival
curves of different groups over time. If these curves maintain a consistent relationship with each other — if they appear to be parallel when plotted on top of one another — it suggests that the proportional hazards assumption may be valid. This is true since parallel survival curves imply that the relative risk or hazard ratios between the groups do not change significantly over the study period, which aligns with the assumption that these ratios are constant over time.

A third method for checking the proportional hazards assumption involves incorporating time-dependent covariates in the Cox model. An interaction between covariates and time that significantly improves model fit could signal a breach of the assumption since the hazards would then clearly not be proportional across the duration of the study.

In summary, validating the proportional hazards assumption involves a multifaceted approach encompassing Schoenfeld residuals analysis, graphical methods, time-dependent covariate inclusion, and formal statistical testing. Ensuring this assumption holds is crucial for the reliability and accuracy of model interpretations.

4. Results

Each of the methods we discussed in the previous section will be utilized in this results section, for more information on the methods you should consult the preceding section. Following the comprehensive evaluation of the Cox Proportional Hazards model's assumptions, we now present our results. The work on the data begins with cleaning. This was not a difficult task but did involve removing several rows containing either all NAs or several NAs. It is noted in the original paper that the rows which I chose to remove, containing only several NAs, were kept for measuring some tasks but not the tasks which contained an NA. Since these individuals were measured differently than others, I felt it was best to either refrain from using the individuals, as I did, or to have imputed their data for the missing tasks. After cleaning, my data contained 22 early dropouts and 103 completers. Also notably, my data cleaning did not account for substance use frequency due to my lack of understanding the original decision boundary of the study. It was stated in the original study that there were 20 early dropouts and 102 completers used in their analysis. Next, I looked into the average time on the PASAT which was 217.31 seconds with a standard deviation of 166.79 seconds. It was noted in the original paper that the average time on the PASAT was 208.7 seconds with a standard deviation of 165.2 seconds.

Next, I chose to use a Welch two sample t-test to compare the PASAT total time between individuals who dropped out before 30 days of treatment and those who completed the treatment. The test did not find a statistically significant difference in the average PASAT Total Time between the dropout group (mean = 173.36 seconds) and the completer group (mean = 226.7 seconds), $t(31.33) = -1.3957$, $p = 0.1726$. It's important to note that the original paper did not use a Welch's two sample t test since it was making a comparison with dysphoria which was not data I was provided. Importantly, the test which I have performed, while significant, was not the same as was performed in the original paper. This lead to different formulas for calculating the degrees of freedom, as was noted in my methods section.

After this, I looked into the other psychological distress task–MTPT-and found the average to be 197.62 seconds with a standard deviation of 94.84 seconds. In the original paper these were found to be 197.1 seconds and 95.9 seconds respectively, however, a correlation test between the PASAT and MTPT measures of psychological distress resulted in a similar conclusion of 0.35 compared to the original paper's 0.38-both with $p < 0.001$. The last descriptive statistics were for BH with an average of 30.13 seconds and a standard deviation of 13.55 seconds and for CPT with an average of 97.26 seconds and a standard deviation of 102.86 seconds. In the paper these were reported as 30.12 seconds, 13.8 sec-

onds, 99.97 seconds, and 104.6 seconds respectively. Running the same correlation test between these two measures of physical distress tolerance as I have previously we obtain a correlation coefficient of 0.26 with $p < 0.01$. For each of these correlation coefficients, I also computed their 95% which came out to 0.18 to 0.49 and 0.09 to 0.42, respectively. From here, I partitioned the physical and psychological measure of distress into a group containing only the early dropouts and a group containing only the completers. I then once again calculated the mean and standard deviation for both and will report these compared side by side in the figures below.

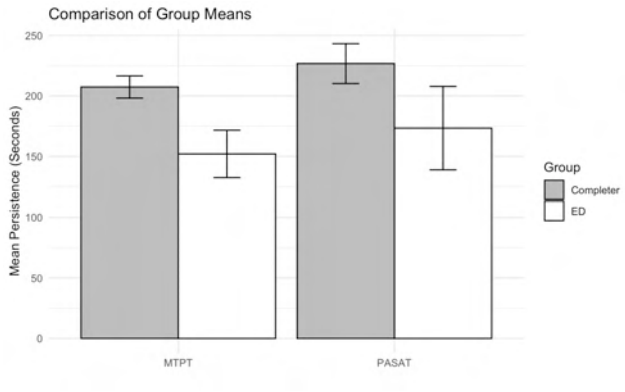


Figure 1: Psychological Distress Measures

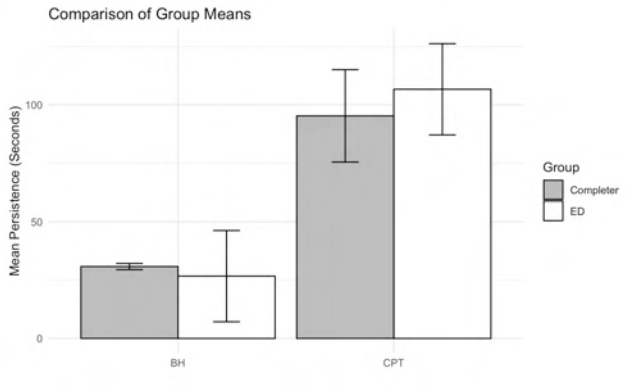


Figure 2: Physical Distress Measures

Notably, there were no group differences between the two physical distress measures while there does appear to be 50 second difference in the group means between the completer and early dropout groups for the two psychological distress measures. This was also found in the paper, although our numbers slightly differ likely due to different sample sizes as noted in the beginning of results.

From here, our analysis proceeded with the application of Cox Proportional Hazards (Cox-PH) models to investigate the relationship between various distress measures on the duration of treatment. I selected three models to consider. The first model incorporated all four distress measures–Breath Holding Duration, Cold Pressor Duration, PASAT Total Time, and Mirror Tracing Total Time. The second model focused on the psychological distress measures-PASAT Total Time and Mirror Tracing Total Time. The third model examined the physical distress measures-Breath Holding Duration and Cold Pressor Duration. Each model was designed to elucidate the distinct and combined impacts of these measures on the treatment duration. In R, the models were written as follows,

```
# Model 1: All Measures
cox.mod <- coxph(Surv(`Days in Treatment`, drop30d) ~
    `Breath Holding Duration` + `Cold Pressor Duration` +
    `PASAT Total Time` + `Mirror Tracing Total Time`, data = datamod)

# Model 2: Psychological Distress Measures
cox.mod2 <- coxph(Surv(`Days in Treatment`, drop30d) ~
    `PASAT Total Time` + `Mirror Tracing Total Time`, data = datamod)

# Model 3: Physical Distress Measures
```

```
cox.mod3 <- coxph(Surv(`Days in Treatment`, drop30d) ~
    `Breath Holding Duration` + `Cold Pressor Duration`, data = datamod)
```

The coxph() function from the survival package in R was used to generate the Cox Proportional Hazards models, and the summary() function in R was utilized to extract the coefficients, standard errors, z-scores, and p-values for each predictor in the models. It is of note, that datamod is our original data set after it has been cleaned and contains both the early dropouts and completers.

From the summary of the first model, we see that the most extreme association is a slightly negative association of Breath Holding Duration with Days in Treatment with coef = -0.024, p = 0.17. Following this, Cold Pressor Duration and PASAT Total Time exhibited non-significant ($p > 0.05$) associations which were both relatively small. Finally, the Mirror Tracing Total Time was found to be significant with coef = -0.005, p = 0.04.

From the summary of the second model, we see that both PASAT Total Time and Mirror Tracing Total Time showed negative associations with the outcome, with MTPT being significant coef = -0.00519, p = 0.03.

From the summary of the third model, we see that both neither 'Breath Holding Duration' nor 'Cold Pressor Duration' showed significant associations with Days in Treatment.

The results from these Cox Proportional Hazards models provide valuable insights into how different distress measures impact expected treatment duration. The significant association of Mirror Tracing Total Time in both the full model and the psychological distress-only model suggests its relevance in predicting treatment outcomes, while the lack of significant findings for physical distress measures might indicate a lesser impact of these measures on treatment duration, or alternatively, the need for a larger sample size with different physical measures to detect an effect.

Next, residual analyses to test the linearity assumption on the second model were conducted using Martingale and Deviance residuals. Below martingale residuals were plotted against the fitted values from Model 2 and a smoothing line was added to aid in the identification of any non-linear patterns. The R code for generating this plot is as follows,

```
plot(predict(cox.mod2), residuals(cox.mod2, type = "martingale"),
    xlab = "fitted values", ylab = "martingale residuals",
     main = "Residual Plot", las = 1)
abline(h = 0)
lines(smooth.spline(predict(cox.mod2),
      residuals(cox.mod2, type = "martingale")), col = "red")
```

Here, a lack of a pattern or trendline in the plot suggests that the linearity assumption is reasonable for the variables in the model.

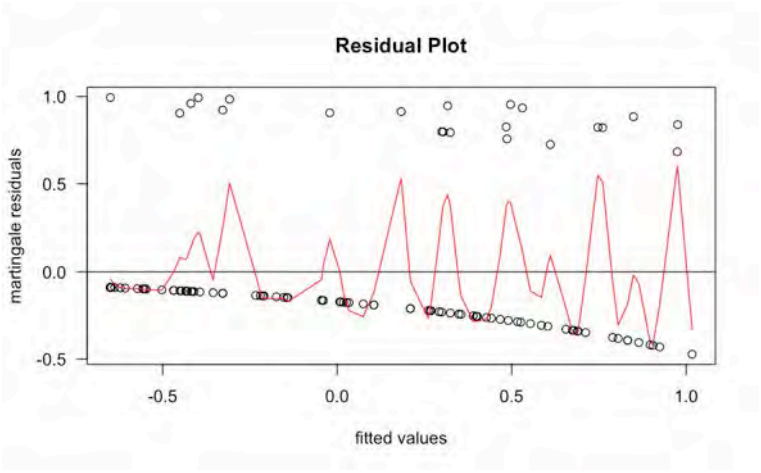


Figure 3: Martingale Residual Plot

Since the red line above clearly lacks any non-linear patterns this plot supports the linearity assumption for the psychological distress measures which were used in the model. Next, deviance residuals were also examined to ensure the robustness of the linearity assumption. Similar to the Martingale residuals, a plot of Deviance residuals against fitted values was constructed, looking for patterns that might indicate violations of the linearity assumption. The R code for generating this plot is as follows,

```
plot(predict(cox.mod2), residuals(cox.mod2, type = "deviance"),
    xlab = "fitted values", ylab = "deviance residuals",
    main = "Residual Plot", las = 1)
abline(h = 0)
lines(smooth.spline(predict(cox.mod2),
    residuals(cox.mod2, type = "deviance")), col = "red")
```

No patterns were identified as the residuals appear to be randomly scattered around the horizontal line at zero, further confirming the linearity assumption for the data.

The final assumption to look into was the proportional hazards assumption which was evaluated using the Schoenfeld residuals. The cox.zph() function from the survival package in R was utilized to perform this assessment.

The Schoenfeld residuals test is used to determine if the effects of predictors are constant over time. A global test and individual tests for each covariate were conducted. The R code used for these tests is as follows:

```
# Global test for the model
cox.zph(cox.mod)
```

The output of this test includes a global chi-squared statistic and individual statistics for each predictor. A non-significant global test suggests that the assumption of proportional hazards is reasonable for the model. In our case the global chi-squared statistics was 5.7, p = 0.06 which is non-significant.

To further test this, individual plots for each covariate in the second model were generated to visually assess the proportional hazards assumption. The R code is as follows,

```
# Time-varying effect of PASAT Total Time
```

```
plot(cox.zph(cox.mod2)[1], xlab = "Transformed Time",
     ylab = "Scaled Schoenfeld Residuals",
     main = "Time-Varying Effect of PASAT Total Time")

# Time-varying effect of Mirror Tracing Total Time
plot(cox.zph(cox.mod2)[2], xlab = "Transformed Time",
     ylab = "Scaled Schoenfeld Residuals",
     main = "Time-Varying Effect of Mirror Tracing Total Time")
```

The plots of the scaled Schoenfeld residuals should show no significant patterns or trends over time if the assumption holds.

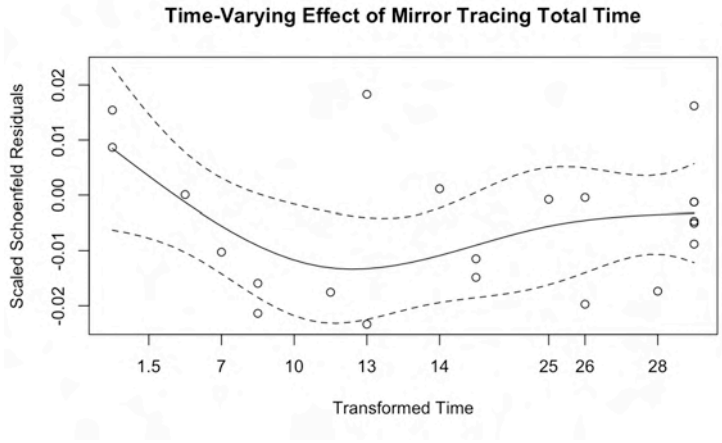


Figure 4

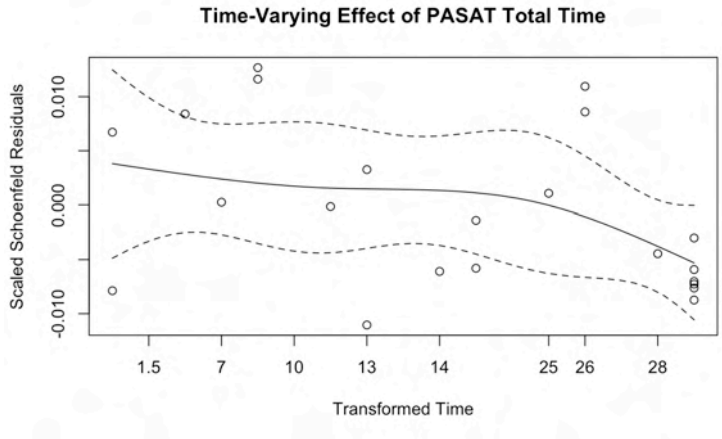


Figure 5

Departures from the horizontal line at $y = 0$ can be indicative of non-proportional hazards, as the assumption of proportional hazards posits that estimates for each covariate should remain relatively constant over time. In the analysis of our second model, there were no significant deviations from this line for either covariate. It is important to consider the scale on the y-axis, which in this case is quite small. This suggests that any minor departures from the horizontal

line are not substantial. Therefore, the tests we've performed for the proportional hazards assumption indicated that the constant hazard ratios for the predictors over time was a reasonable assumption in our models. This supports the use of the Cox Proportional Hazards models for the data and the interpretations we've derived from them.

5. Analysis

This analysis report has demonstrated the utility of Cox Proportional Hazards models in assessing the impact of psychological and physical distress measures on treatment duration. Consistent with the original paper, our findings underscore the more pronounced association of psychological distress measures—particularly the Mirror Tracing Persistency Task (MTPT)—with treatment dropout. A notable contribution of this analysis report is the rigorous testing of the Cox Proportional Hazards model assumptions, which were presumed but not empirically tested in the original work.

Future research on the original paper should consider employing more sophisticated statistical tests such as the Grambsch and Therneau test, which provides a numerical method for analyzing the correlation of scaled Schoenfeld residuals. Moreover, the utilization of the Breslow estimator for handling tied events or expanding the sample size would likely reduce standard deviations and yield a more representative sample, potentially offering broader insights into the replicability of this research. Investigating a variety of physical and psychological distress measures would also be interesting to demonstrate whether the observed results are an artifact of the specific measures of distress tolerance employed or indicative of underlying psychological processes.

Overall, the both the original study and this analysis clearly demonstrate the importance of the Cox Proportional Hazards model and its significance in statistical methodologies, particularly in the context of psychological research.

6. Reflections

My journey through this research project has been as much about statistical understanding as it has been a journey into the intricate world of psychology and addiction. Although my statistical foundation was laid in prior coursework, my understanding of psychology, particularly regarding addiction, was increased substantially through interdisciplinary learning. The seminal work *Unbroken Brain* [Szalavitz, 2016], participation in BRANE lab weekly meetings and lab work, and an extensive repository of shared knowledge through Sharepoint provided a rich education for this study. I also began my dive into discrete event history analysis through this project and have read countless papers on the topic throughout the course of the semester.

This analysis report illuminated the relevance of survival analysis within original research that has garnered significant academic attention, being cited in nearly 400 subsequent papers. Initially, I harbored skepticism about the practical application of the Cox Proportional Hazards model. However, through hands-on application, my appreciation for the semi-parametric nature of the Cox model has grown, especially its flexibility in not requiring a predetermined baseline hazard function.

While it is common in psychology papers to postulate the broader implications of empirical findings, I found it both enlightening and rewarding to delve deeper into the data. Conducting additional tests and analyses beyond which were assumed to be true in the original paper not only strengthened my statistical abilities but also served to bolster the arguments presented by the original author.

Appendix

	PROPORTION (a) IN ONE TAIL								
	.25	.20	.15	.10	.05	.025	.01	.005	.0005
	PROPORTION (a) IN TWO TAILS COMBINED								
df	.50	.40	.30	.20	.10	.05	.02	.01	.001
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.578
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.600
3	0.765	1.978	1.250	1.638	2.353	3.182	4.541	5.841	12.924
4	0.741	1.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.689
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.660
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.460
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.290

Figure 6: T-Score Table [Cote et al., 2021]

References in the Discussion

[Anderson et al., 2023] Anderson, D., Sweeney, D. J., and Williams, T. A. (2023). Statistics. In *Encyclopedia Britannica*. Encyclopedia Britannica, Inc.

[Bradburn et al., 2003] Bradburn, M., Clark, T., Love, S., and Altman, D. (2003). Survival analysis part II: multivariate data analysis–an introduction to concepts and methods. *Br J Cancer*, 89(3):431–436.

[Brown et al., 2002] Brown, R., Lejuez, C., Kahler, C., and Strong, D. (2002). Distress Tolerance and Duration of Past Smoking Cessation Attempts. *Journal of Abnormal Psychology*, 111:180–5.

[Brown et al., 2018] Brown, R., Overstreet, C., Sheerin, C., Berenz, E., Hawn, S., Pickett, T., McDonald, S., Danielson, C., and Amstadter, A. (2018). The Nomological Network of a Behavioral Distress Tolerance Task in Veterans. *J Trauma Stress*, 31(6):876–885.

[Clark et al., 2003] Clark, T., Bradburn, M., Love, S., and Altman, D. (2003). Survival analysis part I: basic concepts and first analyses. *Br J Cancer*, 89(2):232–238.

[Cote et al., 2021] Cote, L. R., Gordon, R. G., Randell, C. E., Schmitt, J., and Marvin, H. (2021). *Introduction to Statistics in the Psychological Sciences*. University of Missouri–St. Louis.

[Cox, 1972] Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220.

[Daughters et al., 2005] Daughters, S. B., Lejuez, C. W., Bornovalova, M. A., Kahler, C. W., Strong, D. R., and Brown, R. A. (2005). Distress tolerance as a predictor of early treatment dropout in a residential substance abuse treatment facility. *J Abnorm Psychol*, 114(4):729–734.

[Ekman, 2017] Ekman, A. (2017). Variable selection for the Cox proportional hazards model: A simulation study comparing the stepwise, lasso and bootstrap approach. Unpublished master's thesis.

[Gillespie, 2006] Gillespie, B. (2006). Checking Assumptions in the Cox Proportional Hazards Regression Model. In *Midwest SAS Users Group (MWSUG) Conference*, Dearborn, Michigan. University of Michigan.

[Goel et al., 2010] Goel, M. K., Khanna, P., and Kishore, J. (2010). Understanding survival analysis: Kaplan-Meier estimate. *International Journal of Ayurveda Research*, 1(4):274–278.

[Herge et al., 2013] Herge, W., Landoll, R., and La Greca, A. (2013). Center for Epidemiologic Studies Depression Scale (CES-D). In Gellman, M. and Turner, J., editors, *Encyclopedia of Behavioral Medicine*. Springer, New York, NY.

[Huang, 2022] Huang, H. (2022). Impulsivity. *Encyclopedia MDPI*.

[Kübler, 2013] Kübler, U. (2013). Structured Clinical Interview for DSM-IV (SCID). In Gellman, M. and Turner, J., editors, *Encyclopedia of Behavioral Medicine*. Springer, New York, NY.

[Patrick and Kramer, 2017] Patrick, C. and Kramer, M. (2017). Multidimensional Personality Questionnaire (MPQ). In Zeigler-Hill, V. and Shackelford, T., editors, *Encyclopedia of Personality and Individual Differences*. Springer, Cham.

[Stewart, 2023] Stewart, K. (2023). Pearson's correlation coefficient. In *Encyclopedia Britannica*. Encyclopedia Britannica, Inc.

[Szalavitz, 2016] Szalavitz, M. (2016). *Unbroken Brain: A Revolutionary New Way of Understanding Addiction*. St. Martin's Press, New York.

[The Editors of Encyclopaedia Britannica, 2023] The Editors of Encyclopaedia Britannica (2023). Student's t-test. Encyclopedia Britannica.

[Tombaugh, 2006] Tombaugh, T. N. (2006). A comprehensive review of the Paced Auditory Serial Addition Test (PASAT). *Archives of Clinical Neuropsychology*, 21(1):53–76.

[Tran, 2013] Tran, V. (2013). Positive Affect Negative Affect Scale (PANAS). In Gellman, M. and Turner, J., editors, *Encyclopedia of Behavioral Medicine*. Springer, New York, NY.

[von Baeyer et al., 2005] von Baeyer, C., Piira, T., Chambers, C., Trapanotto, M., and Zeltzer, L. (2005). Guidelines for the cold pressor task as an experimental pain stimulus for use with children. *J Pain*, 6(4):218–27.

[West, 2021] West, R. (2021). Best practice in statistics: Use the Welch t-test when testing the difference between two groups. *Annals of Clinical Biochemistry*, 58(4):267–269.

Meet the Author:

Claire Heetderks is currently a sophomore at UNC-Chapel Hill studying medical anthropology and exercise and sports science interested in pursuing a career in public health and medicine. Claire first became interested in studying malaria after taking a first-year seminar course focused on the ecology of disease. She continued her independent research within her first-year English course which focused on a research topic of the student's choosing. Claire is also currently an undergraduate research assistant within the Peitrosome Lab Group where she helps with data collection regarding ACL gait retraining post-reconstruction. Claire is passionate about research regarding improving global health through both clinical and anthropological approaches.

Malaria: Difficulties Contributing to Eradication Efforts within Central Africa

Claire Heetderks

Class of 2026

Medical Anthropology BA, Exercise and Sports Science BA

Independent Research

Abstract

This paper explores the complex interactions among climate change, socioeconomic status, and migration, elucidating their collective impact on malaria eradication efforts. Climate change has contributed to the geographical expansion of malaria transmission due to rising temperatures, altered rainfall, and the biological evolution of mosquitoes, complicating control strategies (Pecor et al., 2022). Furthermore, socioeconomic disparities prevalent in Central African communities significantly affect malaria vulnerability as limited access to healthcare, inadequate infrastructure, and educational inequality exacerbate the burden of malaria, disproportionately affecting marginalized populations. Socioeconomic factors intertwine with climate-induced changes as low-income communities can afford to live in areas that are more susceptible to malaria. They also cause increased migration into urban areas, amplifying the susceptibility of specific communities to malaria transmission. Human movement, whether due to economic reasons or environmental changes, can introduce the parasite into new regions, fostering the spread of drug-resistant strains and challenging surveillance and control measures. Overcoming barriers to malaria eradication in Central Africa necessitates a comprehensive, multidisciplinary approach that acknowledges and addresses the intricate connections between climate change, socioeconomic factors, and migration.

Keywords: malaria, climate, socioeconomic status, migration, policy

Malaria: Difficulties Contributing to Eradication Efforts within Central Africa

Introduction

Malaria is a disease spread by mosquitoes that kills around 650,000 people each year (Centers for Disease Control, 2022). Four countries in Africa, Nigeria, the Democratic Republic of Congo, Tanzania, and Niger, account for over half of malaria deaths worldwide (World Health Organization, 2022). While there are still some malaria cases within the United States (US), it was considered eradicated and no longer a public health threat in 1949 (Centers for Disease Control, 2022). Other countries, including those within central Africa, have struggled to eliminate the disease, causing public health officials to ask why. As governments and global aid organizations create new public health policies to implement within these countries, they must consider the factors that influence malaria, including climate, social, and migratory impacts, as these factors most heavily contribute to the spread and severity of the disease.

Historical US Prevention Methods

To understand the state of malaria eradication today, it's crucial to analyze the advantages, disadvantages, and consequences of the tactics used by the United States to eradicate the disease in the 1940s as they give a better picture of what preventative measures can be effective. The United States' eradication of malaria began in 1942 when the Centers for Disease Control's (CDC) predecessor, the Office of Malaria Control, was established to limit the impact of malaria and other vector-borne diseases within military training bases during World War II (Centers for Disease Control, 2020). The US government willingly funded the Office of Malaria Control to ensure the health of troops and medical staff and create new antimalarial drugs that would allow the US army to succeed. First used during World War II,

antimalarial drugs were given to soldiers who traveled to areas with high case rates or known outbreaks of malaria (Centers for Disease Control, 2020).

Along with educating and preserving the military's health, the Office of Malaria Control also created the National Malaria Eradication Program to ensure the health of the civilians. This program, first carried out in the southern US in the 1940s, focused on environmental management consisting of spraying homes with dichlorodiphenyltrichloroethane (DDT), a highly toxic insecticide, as well as water drainage, removal of mosquito breeding sites, and creation of safer homes through the implantation of screened windows and closed roof eaves (Centers for Disease Control, 2020).

The US public was also educated on the threats of malaria with a famous Walt Disney short film. The film portrays the seven dwarves filling in puddles, sleeping with bed nets, and spraying Paris Green, a green insecticide, on their homes and lakes (Walt Disney, 1943). The film was also distributed within Latin American countries as the US wanted to educate and ensure the health of allied soldiers. Countries currently use much safer insecticides than DDT and Paris Green due to their now-known extreme negative consequences on soil, biodiversity, and human health.

The use of insecticides, housing reconstructions, public awareness campaigns, and the United States government's persistence and deep funding allowed the United States to fully eradicate malaria within the country by the end of the 1940s. While these measures worked within the US in the early 1900s, they are challenging to implement in African countries today as most mosquitoes have adapted to be resistant to insecticides, and countries currently lack the funding to implement changes in areas such as housing

reconstruction or public awareness campaigns. While African countries work to develop new preventative measures and policies, it is still essential to implement education on malaria and distribute current preventative measures. Education outreach and distribution of preventative measures still significantly limit case numbers and the severity of infection, as community members report knowing more about malaria and the preventative measures they should be using due to educational outreach (Onyinyechi, 2023).

General Transmission & Current Preventative Methods

Malaria spreads through female mosquitoes, specifically, Anopheles mosquitoes, which are the primary vector of the parasite plasmodium and, in turn, malaria (Centers for Disease Control, 2022). When female mosquitoes go to take blood, they infect humans or animals with plasmodium, spreading the parasites into the blood, where they multiply within the liver and red blood cells (Centers for Disease Control, 2022). If another mosquito goes to drink the blood of the now-infected human, it becomes infected with the plasmodium parasite, and the cycle repeats itself, causing outbreaks. Due to the long life cycle of the parasite within humans, malaria can be mild or severe, with symptoms ranging from a typical virus to more severe cases ending in death (Centers for Disease Control, 2022).

Since malaria is transmitted via mosquito bites, current preventative measures focus on limiting mosquito populations and controlling their interactions with human populations. Current preventative measures include insecticides, wearing long-sleeved clothing, bed nets, and antimalarial medications. Insecticides sprayed on houses work by targeting the insect's nervous system and killing the insect within minutes or hours of direct contact or ingestion of the chemical. Bed nets offer

another layer of protection within homes as the nets prevent mosquitoes from biting when individuals are sleeping. Lastly, antimalarial medications work to kill the malaria parasite during its development stages and prevent the symptoms of the disease (Hill, 2011). Currently, these preventative measures are widely considered the best defense against malaria and must continue to be distributed within high-transmission regions through the help of aid organizations.

Promising New Preventative Methods

To surpass the limitations of current preventative measures, scientists have more recently sought to develop vaccines and new technologies as a long-term public health solution (Ochomo et al., 2022). Current limitations of preventative measures portray that bed nets are becoming less effective as anopheles mosquitoes are becoming resistant to the insecticides sprayed on the nets. In response to these evolving limitations, Dr. Ochomo and other researchers have been experimenting with a new technology, spatial repellents, in Kenya (Ochomo et al., 2022). Spatial repellents are plastic films placed on walls that release chemicals into the air over time and interfere with the mosquitoes' ability to find a host, preventing them from entering homes and biting (Ochomo et al., 2022). Along with the technology, researchers have developed new insecticides and chemicals to use within the spacial repellents that mosquitoes have not encountered or become resistant to. These new chemicals can also be distributed to bed nets to ensure the most successful prevention against mosquitoes.

Along with new preventative measures such as spatial repellents, malaria vaccines are also currently being researched and produced. However, very few malaria vaccines have reached the threshold efficacy rate of 50% (Hill, 2011). Efficacy varies with each disease type and is

measured by how many people who get vaccinated develop the outcome of interest compared with how many who get the placebo and produce the same result. High efficacy rates are challenging to achieve for malaria as the parasites have a complex and adaptable life cycle; the parasites can produce thousands of different antigens, making it difficult for scientists to understand and target the disease (World Health Organization, 2022).

Currently, there are twelve malaria vaccine approaches, the most common of which are blood-stage vaccines, whole parasite vaccines, and mosquito-stage vaccines. Blood-stage vaccines target the disease-causing stage of the plasmodium parasite life cycle, while whole parasite vaccines block the establishment of parasites within the host (Hill, 2011). Lastly, mosquito-stage vaccines target antigens from the plasmodium in the liver stages (Hill, 2011). The current most promising malaria vaccine is RTS, S/AS01, a four-dose vaccine that combines these approaches and entirely acts against *Plasmodium falciparum*, the deadliest malaria parasite (Laurens et al., 2020). Clinical trials have shown significant reduction against malaria, especially severe malaria in children. Over one million children in Ghana, Kenya, and Malawi have been vaccinated. They benefit from the vaccine's added protection as part of a pilot program (Laurens et al., 2020). While the vaccine does not have a 50% efficacy rate, this is the first malaria vaccine recommended by the World Health Organization (WHO) and the European Medicines Agency, two world-renowned public health agencies. The creation of this vaccine for children shows a step in the right direction toward future preventative methods for malaria.

Climate Change

Climate influences a country's ability to stop the spread of malaria as

mosquitoes thrive in warm, wet environments, and climate change has increased the scope of their possible environment. Decreasing yearly rainfall attributed to climate change allows mosquitoes to breed longer as the rainfall no longer wipes out the stagnant water in which mosquitoes breed. This decrease in moisture caused by climate change has also caused an increase in urbanization within Africa as subsistence agriculture is no longer dependable due to droughts. Hence, people migrate to urban areas for better economic opportunities (Pecor et al., 2022). As communities migrate to urban areas, population density maybe use the word directly? increases, leading to more accessible proximities for anopheles mosquitoes to infect individuals and pass malaria on to one another. Along with decreasing rainfall, higher temperatures due to climate change also influence the spread of malaria as it allows the mosquito's environment to expand geographically. Specifically within central Africa, areas of high altitude are now experiencing malaria outbreaks as temperature increases have allowed mosquitoes to breed in areas that were initially too cold for mosquitoes to survive. (Vilenna et al., 2020).

One study found that the temperature at which anopheles mosquitoes can transmit disease has increased, allowing the months of transmission and regions of transmission within Africa to widen in scope (Vilenna et al., 2020). This now enables anopheles mosquitoes to breed almost all months of the year and in urban areas where they initially could not survive. The increased difficulty of eradicating malaria due to climate change is apparent with a new type of anopheles mosquito in Djibouti, *Anopheles Stephensi* (Pecor et al., 2022). First reported in Djoubti in 2012, *Anopheles Stephensi* arrived from Asia and is resistant to all insecticides. It has

also biologically adapted to live in urban areas and survive in dry seasons, allowing the species to have a much vaster range of possible infections (Pecor et al., 2022). As communities migrate to urban areas for better economic opportunities, individuals are at higher risk of getting much sicker from malaria as they don't have acquired immunity from previous exposure. One research study has concluded that the percentage of *Anopheles stephensi* has increased in regions outside Djouboti, where the spread began (Pecor et al., 2022). For example, Sudan, a country with previously low malaria rates, experienced a malaria epidemic in 2012 due to *Anopheles Stephensi* (Abubakr et al., 2022). As yearly temperature and rainfall continue to vary due to climate change, the environment of mosquitoes will continue to adapt and evolve, allowing malaria outbreaks to continue without proper intervention.

Socioeconomic Status

Socioeconomic status influences a country's ability to eradicate malaria, as many poverty-stricken countries and regions are more prone to mosquitoes and are not given the resources or financial help needed to prevent infection. Countries that lack resources such as bed nets, insecticides, and rapid treatment of antibiotics are more prone to outbreaks that spread faster (Idris et al., 2022). Even countries with resources are influenced by corruption and bribery within the government and health systems, as bribes are sometimes used to get care faster. Poorer patients are more reliant on public services and, therefore, more vulnerable to bribery (Hsiao et al., 2019). Patients are confronted with having to pay the bribe or delaying seeking care until they are much sicker. Corruption also affects communities because distrust of the healthcare system can influence whether individuals seek treatment. Within a case study reviewing the malaria outbreak in South Sudan in

2018, the authors assessed clinic data collected from participants who had received a recurrent malaria diagnosis and had visited one of the three primary healthcare centers. This research provides evidence that when patients don't complete their prescribed treatment regimen, the severity of malaria recurrence drastically increases as the few parasites that survive due to not completing the treatment regimen become resistant to the drugs used. The authors concluded that factors such as marital status, employment status, use of preventative measures, and nutrition status all affected how willing participants were to take the antibiotics (Idris et al., 2022). For example, within this case study, the authors concluded that some individuals were not using their bed nets properly as a preventative technique and instead used them for fishing. When interviewed, the individuals cited that they knew of the consequences but were considering more pressing matters, such as feeding their families.

Socioeconomic status can also influence how malaria is spread due to the relationship between poverty and housing. The risk of malaria infection increases with structures that contain cracks or openings because they open up more space for mosquitoes to enter homes and infect individuals. These homes are often found in poverty-stricken areas as these individuals cannot afford other housing. These homes are also usually located close to ideal mosquito breeding grounds as these areas are often the least expensive and desirable for community development (Idris et al., 2022). Given this research, when considering what best resources and techniques should be used for malaria eradication, differences in socioeconomic status must be regarded to ensure equitable protection from the disease.

Migration

Migration can influence the eradication of malaria as migrants can

reintroduce malaria to areas of low transmission, and inadequate education on malaria can cause a lack of surveillance and accountability. Some individuals may have malaria and be unaware, causing outbreaks in the new areas they migrate to. Migrants traveling from low to high-transmission regions are more likely to acquire severe malaria as they have not built up acquired immunity (Lin et al., 2022). In comparison, migrants who travel to low-transmission areas or communities with high immunity due to a past outbreak may reintroduce malaria. Farmers commonly spread malaria across borders as some will spend multiple days within a farming site before traveling back to be with their families when not working. This spreads malaria to new populations or reintroduces malaria into a low-transmission area.

In one case study on migration patterns, researchers tested hundreds of workers via PCR tests within the work sites. They concluded that a majority had contracted malaria through their border work. Through the local public health agency, workers took antibiotics and bed nets and received an educational lesson on the spread of malaria. The camp where workers stayed was treated with insecticides over the next several days. The local government and CDC followed up with patients days or weeks later to ensure the patients were keeping up with their treatment and understood how the disease could spread through their community (Lin et al., 2022). This case study and other research suggest that surveillance of migrants and malaria patients remains successful in helping prevent the spread of malaria. Given the positive impact of the preventative measures used within this case study, countries should work to distribute the resources needed to avoid malaria and implement effective surveillance and response systems to help prevent the spread of malaria caused by migration. While many researchers know this task is daunting, they

believe it is feasible, mainly as malaria migrates into urban areas and areas of higher socioeconomic standing, forcing private corporations and aid organizations to take the problem more seriously.

Health Policy

Malaria policy must be executed locally as local communities are more aware of what resources would benefit their health and communities' well-being (Lin et al., 2022). Community members can ensure that local culture, social norms, and stigma are considered when implementing health policy, which leads to more effective and equitable healthcare delivery. The World Health Organization wants more funding for malaria prevention education, resources, and vaccine research (World Health Organization, 2022). The WHO made ample progress toward reaching these policy milestones until COVID-19 set them off course, as resources and medical staff were incredibly limited, and funding was redirected (World Health Organization, 2022). With public health coming to the forefront of attention due to COVID-19, organizations are reevaluating their goals and ensuring that countries, such as those in central Africa drastically impacted by COVID-19, are set on the right path.

An example of effective health policy change is the WHO's response to the Ebola epidemic in West Africa in 2018. The WHO created a new policy system beginning with the implementation of resources, education, and research at the local level. The WHO also ensured corruption and bribery were considered within their new policy by increasing transparency and getting more local participation within government and healthcare. Research has shown that this policy has been highly effective and allowed for more efficient care and prevention. While this is a step in the right direction, countries that contribute to the funding of the World Health Organization

must hold the WHO accountable and ensure progressive policy is implemented over the next several years.

Conclusion

The response to malaria from central African countries has been different from that of the US in the 1940s due to factors such as, the evolution of mosquitoes, increasing climate change, the economy and healthcare corruption of African countries, and patterns of migration. Health policy should focus on implementing new preventative measures such as newly developed insecticides and spatial repellents, funding and implementing malaria vaccines, and educating the public on malaria. Policy should also focus on longer-term goals of creating better housing infrastructure and malaria surveillance, which would help to prevent future outbreaks. To maximize effectiveness within new health policy, policymakers must enact local cultural practices, opinions, and reasoning. With these new implementations, lives can be saved as new preventative measures and education reduce case numbers and deaths, and eradicating malaria within central Africa and other regions with high transmission can become a practical and attainable goal with success measured through WHO measurements of disease.

References

Abubakr, M., Sami, H., Mahdi, I., Altahir, O., Abdelbagi, H., Mohamed, N. S., & Ahmed, A. (2022). The Phylodynamic and Spread of the Invasive Asian Malaria Vectors, *Anopheles stephensi*, in Sudan. *Biology*, *11*(3), 409. https://doi.org/10.3390/biology11030409

Centers for Disease Control and Prevention. (2022, August 19). *CDC - Parasites - Malaria*. Centers for Disease Control and Prevention. Retrieved February 13, 2023, from https://www.cdc.gov/parasites/malaria/index.html

Hill, A. (2011, October 12). *Vaccines against malaria*. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences. Retrieved April 29, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3146776/

Hsiao, A., Vogt, V., & Quentin, W. (2019, August 21). *Effect of corruption on perceived difficulties in healthcare access in sub-Saharan Africa*. https://www.ncbi.nlm.nih.gov/. Retrieved April 29, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6703670/

Idris, I., Ayeni, G., Iyamu, I., Sina-Odunsi, A., Adebisi, Y., & Obwoya, J. (2022). Factors influencing the severity of recurrent malaria in a conflict-affected state of South Sudan: An unmatched case-control study. *Conflict & Health*, *16*(1), 1–10. https://doi.org/10.1186/s13031-022-00463-z

Laurens, M. B. (2020, March 3). *RTS,S/AS01 vaccine (Mosquirix™): An overview*. Human vaccines & immunotherapeutics. Retrieved April 22, 2023, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7227679/

Lin, Z. R., Yin, S. S., Yang, J., Guo, X. R., Dong, C. L., Lin, Y. K., Ding, C. L., Sun, X. D., Yan, R. X., Yang, S. L., Zhou, X. H., & Xu, J. W. (2022). The public health response to an outbreak of border-spill malaria along China-Myanmar border. *PloS one*, *17*(12), e0275932. https://doi.org/10.1371/journal.pone.0275932

Ochomo, E.O., Gimnig, J.E., Bhattarai, A. *et al.* Evaluation of the protective efficacy of a spatial repellent to reduce malaria incidence in children in western Kenya compared to placebo: study protocol for a cluster-randomized double-blinded control trial (the AEGIS program). *Trials* 23, 260 (2022). https://doi.org/10.1186/s13063-022-06196-x

Onyinyechi, O. M., Mohd Nazan, A. I. N., & Ismail, S. (2023). Effectiveness of health education interventions to improve malaria knowledge and insecticide-treated nets usage among populations of sub-Saharan Africa: systematic review and meta-analysis. *Frontiers in public health*, *11*,

1217052. https://doi.org/10.3389/fpubh.2023.1217052

Pecor, D. B., Potter, A. M., & Linton, Y.-M. (2023, September 29). *Implications of climate change and Anopheles Stephensi Liston in Africa: Knowledge gaps and lessons from history - current Tropical Medicine reports*. SpringerLink. https://link.springer.com/article/10.1007/s40475-023-00296-7

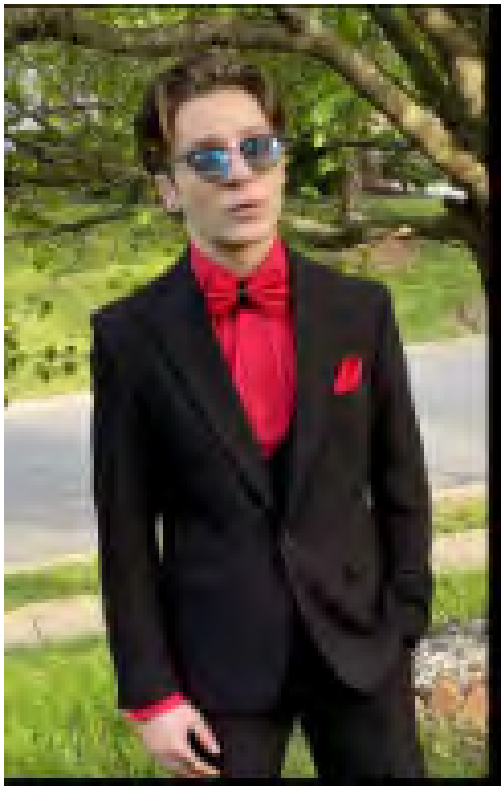
Pekar, J., Magee, A., & Parker, E. (2022, August 26). *The molecular epidemiology of multiple zoonotic origins of SARS-COV-2*. Science (New York, N.Y.). Retrieved April 29, 2023, from https://pubmed.ncbi.nlm.nih.gov/35881005/

Villena, O. C., Ryan, S. J., Murdock, C. C., & Johnson, L. R. (2022, March 22). *Temperature impacts the environmental suitability for malaria transmission by Anopheles gambiae and Anopheles stephensi*. https://esajournals.onlinelibrary.wiley.com/. https://esajournals.onlinelibrary.wiley.com/doi/10.1002/ecy.3685

Walt Disney. (1943). *The Winged Scrouge*. *YouTube*. United States. Retrieved April 22, 2023, from https://www.youtube.com/watch?v=gc4a8bZxCBA.

World Health Organization. (2022, December 8). *Fact sheet about malaria*. World Health Organization. Retrieved February 13, 2023, from https://www.who.int/news-room/fact-sheets/detail/malaria

Meet the Author:



Corey Reichert was raised in North Carolina where they obtained their associate degree in the arts at the age of 17 at Rowan-Cabarrus Community College. While there, he maintained excellent grades and finished with the honor of Summa Cum Laude. He then transferred to the University of North Carolina at Chapel Hill to pursue a bachelor's degree in the majors of political science and peace war and defense. During his first year at UNC-CH, he authored the following essay 'Agenda 2063: The Enforcement of Soft Law' which discusses the failures of the agenda in its implementation and the need for reform so that the document may enforce the Africa that its inhabitants want. He enjoys drawing and reading in his free time.

Agenda 2063:

The Enforcement of Soft Law

Corey Reichert

Policy and Practice Review Article

Academic Year 2025 Political Studies Major

October 22, 2023,

Abstract

The future of the African continent and the subsequent rise is in the fate of a singular plan. That plan is the *Agenda 2063: The Africa We Want*. The project itself is packed with multiple initiatives that would assist and plan for the blossoming of the African continent into a unified geopolitical superpower. Some of the projects are as ambitious as the Silencing the Guns in Africa by 2020, which calls for the ending of conflict on the continent by 2020. Though seeing that 2020 has already passed and the African continent is still rife with conflict, it is effortless to notice that the agenda is gradually failing. While the agenda's current role as a soft law initiator is doing tremendous work, this limit allows remarkably central pursuits and objectives to not be implemented. Though all hope is not lost, there have been numerous cases in which soft law instruments have developed into genuine hard law that is thoroughly enforced. By using prior examples and changes within the system of the African Union's main organizations, such as the Pan-African Parliament, Agenda 2063 could potentially rise into the document that can develop Africa into the future.

Key Words: Agenda 2063: *The Africa We Want,* African Union, Pan-Africanism, African Continental Free Trade Area (AfCFTA), Responsibility to Protect (R2P).

Introduction

The modern world is of global powers and their geo-political influence. The most famous example of this is the current ongoing conflict of influence between a rising Chinese dragon and an established American eagle, both of which wish to hold the title of world hegemon. Though, as with most things, this world isn't bipolar. According to the *"World Population Review,"* the current superpowers on Earth are "The United States, China, European Union, India, and Russia".[1] While these are all undisputed major players within the global stage, there is a notable scarcity of inclusion from nations of the African continent, or just the inclusion of the African Union since it is an influential actor on the global stage and there is a prior inclusion of an entire continent with the European Union. This relegation of Africa as an uninvolved participant on the global stage and just as a battleground of influence leaves it in the position it held during the colonial era. This appointment is one the African Union wishes to break with its implementation of *Agenda 2063: The Africa We Want*. The agenda is a blueprint that will establish Pan-African unity across the continent and allow for the combined influence of the continent to make the continent a global Superpower, similar to the situation of the European Union. This is shown within the African Union's description of the document stating, "Africa's blueprint and master plan for transforming Africa into the global

[1] World Superpowers 2023, accessed November 20, 2023, https://worldpopulationreview.com/country-rankings/world-superpowers.

powerhouse…its goal for inclusive and sustainable development and is a concrete manifestation of the Pan-African drive for unity".[2] This document will bring hope for the future of the African continent into a place of peace and prosperity. However, that hope must be implemented. The lack of institutional power within African Union organizations has been prevalent as multiple instrumental program deadlines have already gone by. Most notable is the 'Silencing the Guns by 2020' which was a program to help make Africa a conflict-free area but has now been "working towards a better definition of what 'silencing the guns' means in the context of the Agenda 2063 goals and specific milestones that should be achieved between now and 2030".[3] Agenda 2063, titled "The Africa We Want", is a landmark blueprint to transform the continent into a global player. The document is a champion of soft law and has been stated as "the core documents of present-day and future African regional integration".[4] However, Its relegation to this role has left it vulnerable to failure as present within other international organizations such as the presence of converging national interest, underwhelming participation, overreliance on international norms, and lack of institutional development.

A. Description of Agenda 2063

Before persisting with the argument at hand, it is necessary to define and understand the inner workings and programs of Agenda 2063. The implementation of the agenda is planned to come in ten-year waves with the first planned implementation lasting from 2013-2023. Though there are multiple implementation plans, the most relevant one

[2] "Agenda 2063: The Africa We Want.," Agenda 2063: The Africa We Want. | African Union, January 1, 2019, https://au.int/en/agenda2063/overview.

[3] PSC Report, "Staying on Target to Silence the Guns by 2030," ISS Africa, June 22, 2022, https://issafrica.org/pscreport/psc-insights/staying-on-target-to-silence-the-guns-by-2030.

[4] Markus Kaltenborn, "Implementing International Social Protection Initiatives in Africa – the Role of Global and Regional Soft Law," *Recht in Afrika* 20, no. 1 (2017): 3–10, https://doi.org/10.5771/2363-6270-2017-1-3, pp. 9.

is the current one as the purpose of it is to set a precedent for future ten-year plans and to see the shortcomings of the agenda. This is shown by the fact that "the African Peer Review Mechanism has been given a wider role …with monitoring and evaluation role for the AU Agenda 2063 and the United Nations Sustainable Development Goals (SDGs) Agenda 2030" and that one of the plan's purposes is to "Identify priority areas, set specific targets, define strategies and policy measures required to implement the FTYIP".[5] [6] The enactment of measures that allow the agenda to be more fluid in execution will help with the potential defects that are currently plaguing the plan. Not only that but shortcomings within previous plans can become the focus of forthcoming plans, thus making no certain aspiration fall behind.

The main goals that direct Agenda 2063 are its seven different aspirations for the continent and the sub-goals that follow those aspirations. Those aspirations are sustainable development, an integrated continent based on Pan-Africanism, good governance, peace and safety, common values, equality, and empowerment of all spheres of life, and an influential Africa on the global stage.[7] These aspirations are exceptionally primary in conception as they are values and goals that are wanted across the world. This is shown by the fact that the African Union has already overlaid the goals present within Agenda 2063 with goals present in the United Nation's Sustainable Development Goals and how they overlap.[8] [9] This overlap of both global and continental obligations, as all 193 countries did sign the SDGs, should be the

[5] Amani, "Mapping of AU Decision Making Actors and Processes," Amania Africa Report, April 2022, https://amaniafrica-et.org/wp-content/uploads/2022/04/Mapping-of-AU-decision-making-actors-and-processes.pdf. pp. 27.

[6] "The First-Ten Year Implementation Plan," The First-Ten Year Implementation Plan | African Union, February 10, 2022, https://au.int/en/agenda2063/ftyip.

[7] "Our Aspirations for the Africa We Want," Our Aspirations for the Africa We Want | African Union, February 10, 2022, https://au.int/en/agenda2063/aspirations.

[8] "Linking Agenda 2063 and the Sdgs," Linking Agenda 2063 and the SDGs | African Union, February 10, 2022, https://au.int/agenda2063/sdgs.

[9] Franck Kuwonu, "Agenda 2063 Is in Harmony with Sdgs | Africa Renewal," United Nations, December 2015, https://www.un.org/africarenewal/magazine/december-2015/agenda-2063-harmony-sdgs.

African Union as it shows to both the world and Africa that its leaders are dedicated to similar goals and the general development of humanity.[10] However, when looking at the developments towards achieving these goals, there has been a significant deficiency in action. Not a single key project of Agenda 2063 has met completion, though some have had more development such as the African Continental Free Trade Area. The project is about destroying the trade barriers that currently separate the continent and allowing for the increase of continental-wide trade through this free trade agreement.[11] This project in particular has had one of the best ratification rates out of all of the other flagship projects with "54 AU Member States have signed the AfCFTA Agreement, 42 Member States have ratified it".[12] This by itself is a monumental movement towards the economic integration of the continent and the progress and political motivation made is a benchmark for what Agenda 2063 can do if properly executed. However, the rest of the '*Second Continental Report on The Implementation of Agenda 2063'* doesn't show the same enthusiasm or progress of enforcement. Flagship projects such as the AfCFTA agreement require little capital investment and political motivation but in return gain economic connections and thus have been much more successful in their ratification, though other programs may infringe on the sovereignty of a nation and thus "have registered slow progress".[13] One notable project that has failed to meet its deadline was that of Silencing the Guns in Africa by 2020. This project, while having made some progress on funding with the AU Peace Fund gaining 77% of the required

10 "Historic New Sustainable Development Agenda Unanimously Adopted by 193 UN Members," United Nations, September 25, 2015, https://www.un.org/sustainabledevelopment/blog/2015/09/historic-new-sustainable-development-agenda-unanimously-adopted-by-193-un-members/.

11 "African Continental Free Trade Area (AfCFTA): Auda," NEPAD, accessed November 20, 2023, https://www.nepad.org/agenda-2063/flagship-project/african-continental-free-trade-area-afcfta.

12 "Second Continental REPORT," African Union, February 2022, https://au.int/sites/default/files/documents/41480-doc-2nd_Continental_Progress_Report_on_Agenda_2063_English.pdf. Pp. 52.

13 "Second Continental REPORT," African Union, February 2022, Pp. 59.

contributions, has failed to end all conflicts within Africa, and fourteen still ravage the continent.[14] [15] As said before, the APRM along with the Pan-African Parliament (PAP) and the African Multidimensional Regional Integration Index (AMRII) are all used to monitor the progress of the flagship projects.[16] The PAP in particular can draft and advise towards future implementation of the agenda.[17] With all of this administrative authorization, the true flaws of Agenda 2063 can be detected. That is the fundamental problem of initiative and enforcement. The agenda's relegation to essentially a guideline for the future, will thus leave it to the responsibility of the willing to enact change. The willing can only do so much for an initiative that requires numerous government bodies to act. If other nations don't ratify or finance the projects that the Agenda 2063 outlines, then the impacts of the aspirations will be meager at best because of the lack of signatures and ratifications on flagship projects. Thus, one of the key issues with Agenda 2063 is its relegation to simply an outline and not something that all African nations are obligated to achieve.

B. Transformation From Soft Law towards Hard Law Through Norm Creation

That obligation could be from Agenda 2063's potential transformation from a soft law instrument to one that would have the authority to enforce its initiatives across national lines as a hard law. This transition would be overambitious and would match the overambitious nature of Agenda 2063. The difference between a hard law and a soft law

14 "Second Continental REPORT," African Union, February 2022, Pp. 3.
15 "Second Continental REPORT," African Union, February 2022, Pp. 151.
16 "Au Rolls out Framework for Monitoring and Evaluating the Status of African Regional Integration," African Union, November 20, 2023, https://au.int/en/articles/au-rolls-out-framework-monitoring-and-evaluating-status-african-regional-integration.
17 "Agenda 2063 - Report of the Commission on the African Union Agenda 2063," African Union, January 31, 2015, https://portal.africa-union.org/DVD/Documents/DOC-AU-WD/Assembly%20AU%205%20(XXIV)%20_E.pdf. pp.13

is that of enforcement. This is shown by the definition of soft law which states, "Co-operation based on instruments that are not legally binding, or whose binding force is somewhat "weaker" than that of traditional law".[18] Though soft law and hard law can be multiple different things, they simply must influence the implementation of policy creation in either binding or non-binding ways. The transition can also be remarkably fluid rather than the conflict and debate that might come to mind when trying to make something that is non-binding to become binding. An example of this would be the ratification of AfCFTA which shows one form of soft law transition. That transition is norm creation, where the influence of Agenda 2063 allowed for the transition of the AfCFTA from simply part of a blueprint to a binding agreement enforced by every state that ratified it. The AfCFTA will not only help with trade but also regional integration. This is shown by its primary goals which state, "By removing tariffs…member states intend to: facilitate intra-African trade; promote regional value chains to foster the integration of the African continent".[19] If this project succeeds then it does have the potential to influence African actors to implement further flagship projects in the hopes that it continues the precedent of success that came with the AfCFTA. However, the main problem that comes from this form of law transition is speed and enforcement. The AfCFTA only advanced as fast as it did because of the "net gains overall" even though "there will be winners and losers".[20] The reason for this is due to the economic uncertainty that comes with trade agreements, usually, the impacts can only be hypothesized and will only be realized upon completion. With this, it is in nearly every

[18] Soft law - OECD, accessed November 20, 2023, https://www.oecd.org/gov/regulatory-policy/irc10.htm.

[19] Andrea Cofelice, "African Continental Free Trade Area: Opportunities and Challenges," *The Federalist Debate* 31, no. 3 (2018): 32–35, https://doi.org/10.2478/tfd-2018-0032. Pp. 32.

[20] Kwabena Nyarko Otoo, "Africa's Economic Trade-Off," Africa's economic trade-off – Economy and ecology | IPS Journal, October 18, 2021, https://www.ips-journal.eu/topics/economy-and-ecology/african-continental-free-trade-area-5496/.

nation's vested interest to be part of the agreement as it is the world's largest trade agreement with fifty-four nations and 1.3 billion people even if it's a gamble.[21] Thus, access to this market of this scale would be beneficial to any nation taking part. But taking away the obvious benefits and the fact that there has already been a precedent of trade agreements within Africa with the multiple economic blocs present on the continent, and the speed which was shown in this initiative would have been just as stalled by political convolution as the rest of the flagship initiatives.[22] Thus, norm creation wouldn't be the most efficient form of policy creation for the agenda mainly because there are specific deadlines and goals that each initiative is supposed to hit. While it can be effective at times, its need for willing participants doesn't coincide with the fact that it would require multiple African leaders whose views will be more selfish towards the immediate prosperity of their nation than that of the entire continent, to agree towards the same details for potentially hundreds of projects that could come with Agenda 2063 while having a ten-year deadline.

C. The transition from Soft Law to Hard Law Through International Measures

International influence on the African continent is no surprise. From direct control by colonialism to the neo-colonialistic agenda present within neo-liberalism, there has been and will always be an outside influence on the continent. Yet, the influence can be one for good and not for greed. This has already been mentioned through the international measures that help obligate the African government to the aspirations of

[21]Maryla Maliszewska, "The African Continental Free Trade Area: Economic and Distributional Effects," World Bank Group, accessed November 20, 2023, https://www.wto.org/english/thewto_e/acc_e/afcfta_feb11maryla.pdf. Pp. 2.

[22]Regional Economic Communities," United Nations Economic Commission for Africa, accessed November 20, 2023, https://archive.uneca.org/oria/pages/regional-economic-communities.

Agenda 2063 through the international obligation to complete the SDGs. A more prominent example of a shift of international soft law into a hard law within the African Union would be that of Article 4(h) and the Constitutive Act of the African Union. This article provides the African Union the right to "intervene in a Member State pursuant to a decision of the Assembly in respect of grave circumstances, namely: war crimes, genocide and crimes against humanity".[23] While this right gives the African Union incredible power to protect principles of human rights through intervention, it was not originally constructed by the African Union but rather one made by the United Nations Secretary-General, Kofi Annan. In his annual report, he calls for "a real and sustained commitment to help end their cycles of violence, and launch them on a safe passage to prosperity".[24] This isn't a direct guideline for any future implementation, but simply a call to action towards future and ongoing atrocities in the world. This call to arms would be answered by international actors such as Canadian Prime Minister, Jean Chrétien and reports such as the International Commission on Intervention and State Sovereignty's report.[25] These soft laws would coagulate into the modern Responsibility to Protect doctrine (R2P), which formed a guideline for intervention in future atrocities.[26] This doctrine is still a soft law mechanism but, by no means, does any nation within the UN have to follow it or abide by it. It's simply a guide in case of future atrocities. Yet, in Africa, that has changed. By some bizarre chance, the early 2000s was also a change for

[23]Ben Kioko, "The Right of Intervention under the African Union's Constitutive Act: From Non-Interference to Non-Intervention," *International Review of the Red Cross* 85, no. 852 (2003): pp. 807, https://doi.org/10.1017/s0035336100179948.

[24]Kofi Annan, "Secretary-General Presents His Annual Report to General Assembly | UN Press," United Nations, September 20, 1999, https://press.un.org/en/1999/19990920.sgsm7136.html.

[25]Ololade Shyllon and Busingye Kabumba, "Chapter 9: Soft Law and Legitimacy in the African Union: The Case of the Pretoria Principles on Ending Mass Atrocities Pursuant to Article 4(H) of the AU Constitutive Act," essay, in *The Model Law on Access of Information for Africa and Other Regional Instruments: Soft Law and Human Rights in Africa* (Pretoria: Pretoria University Law Press, 2018). Pp 178-180.

[26]Augustin Hodali, THE IMPLEMENTATION OF THE RESPONSIBILITY TO PROTECT (R2P) NORMS BY THE AFRICAN STANDBY FORCE IN SUB-SAHARAN AFRICA, June 9, 2017, https://apps.dtic.mil/sti/pdfs/AD1038733.pdf. Pp. 1.

the Organisation of African Unity into the African Union.[27] Any institutional changes leave an opportunity to implement new ideas, which happened to be the R2P doctrine. This would be especially critical for Africa as it just suffered through the Rwandan Genocide and the ongoing Congo Wars. It had lived and was continuing to live through the failures of the UN's inability to enforce the soft law which was R2P. Thus, it only made sense for Africa to enforce it. This came with the previously mentioned Article 4(h). This process of an international figure influencing the creation of a doctrine for a global organization was then used to empower the African continent through the hard law of its constitution is nothing less than an incredible example of how soft law can be put into effect. However, this outcome isn't all acceptable. The main problem stems from having international influence create such momentous laws is that they are thus relegated to follow the soft law that came from it. This came in the form of the Pretoria Principles. The Pretoria Principles was a conference of policymakers and academics that was set to enhance the African Union's role in article 4(h) as it has never used it before.[28] The lack of legitimacy that plagued the conference due to the lack of authority on the matter and lack of public participation which thus forced it to seek other means of legitimacy.[29] This was with the form of coherence with other forms of a doctrine similar to Article 4(h), that being the R2P doctrine. Thus, to gain legitimacy, the Pretoria Principles called for "the AU requires the authorization of the UN Security

[27]"About the African Union," About the African Union | African Union, February 10, 2022, https://au.int/en/overview.

[28]Ololade Shyllon and Busingye Kabumba, "Chapter 9: Soft Law and Legitimacy in the African Union: The Case of the Pretoria Principles on Ending Mass Atrocities Pursuant to Article 4(H) of the AU Constitutive Act," essay, in *The Model Law on Access of Information for Africa and Other Regional Instruments: Soft Law and Human Rights in Africa* (Pretoria: Pretoria University Law Press, 2018). Pp 181-182.

[29]Ololade Shyllon and Busingye Kabumba, "Chapter 9: Soft Law and Legitimacy in the African Union: The Case of the Pretoria Principles on Ending Mass Atrocities Pursuant to Article 4(H) of the AU Constitutive Act," essay, in *The Model Law on Access of Information for Africa and Other Regional Instruments: Soft Law and Human Rights in Africa* (Pretoria: Pretoria University Law Press, 2018). Pp 182.

Council for article 4(h) intervention' and that 'the UN Security Council has the responsibility to authorize the use of force in the implementation of article 4(h) intervention".[30] The requirement of UN Security Council authorization thus diminished the power that was previously exclusively held by the AU. Thus, in a bid to overlap the article with international norms, it has eradicated the element of African agency which was initially championed in Article 4(h). While international influence is valuable and can help boost the development of norms and laws, it can be used against the wishes of the African Union such as R2P has. This arrangement would be unacceptable for something as pivotal as Agenda 2063. This blueprint is for "*The Africa We Want*" not the Africa the world wants.[31]

D. Institutional Transition from Soft to Hard Law

Previous sections have delved into individual cases of soft law transition into hard law, though this is fitter towards the truth of its transition. Without true institutional organisms or radical change as seen with Article 4(h), soft law will only become enforced when conditions permit it to. While this can still lead to change and further integration towards Pan-Africanism, it would do it at a much slower rate than necessary for the strict deadlines of Agenda 2063. This thus makes the requirement of institutional changes to help with pursuing the soft-law planning of Agenda 2063 into hard-law implementation. A streamlined process for the strict deadlines of the various aspirations and flagship projects. The institution in particular that could have the

[30]Ololade Shyllon and Busingye Kabumba, "Chapter 9: Soft Law and Legitimacy in the African Union: The Case of the Pretoria Principles on Ending Mass Atrocities Pursuant to Article 4(H) of the AU Constitutive Act," essay, in *The Model Law on Access of Information for Africa and Other Regional Instruments: Soft Law and Human Rights in Africa* (Pretoria: Pretoria University Law Press, 2018). Pp 184.

[31]"Agenda 2063: The Africa We Want.," Agenda 2063: The Africa We Want. | African Union, January 1, 2019, https://au.int/en/agenda2063/overview.

potential to fill this role is the Pan-African Parliament. The reason this organizational body is above the rest is due to the administrative deficits it already suffers from and its current role as a monitor of Agenda 2063. The former reason is shown by the fact the PAP has never passed a single law and "It is the combination of its deliberative and recommendation power and significantly the power to propose model laws that give PAP its quasi-legislative role, as such model laws constitute a soft-law instrument of the AU".[32] [33] However, before giving authority to any organization it is necessary to see if it has the legitimacy to be given it. Concerning the PAP, it excels in one form of legitimacy, in particular, that is democratic participation. The importance of this is that the inclusion of both non-state actors and state actors allows more people to believe and work in the system thus building a form of trust in its operations.[34] This trust gives the PAP legitimacy as the entire point of the parliament as laid out by Article 17(1) of the Constitutive Act which states, "the Parliament is intended as a platform for people from all African states to be involved in discussions and decision-making on the problems and challenges facing the continent".[35] Thus enshrined within the document that birthed the parliament is the motive for its use to be universal around the continent and for the furtherment of Pan-Africanism pursuits which not only give the parliament its name but also align it with things such as Agenda 2063. However, the

[32]Babatunde Fagbayibo, "Toothless Pan-African Parliament Could Have Meaningful Powers. Here's How," The Conversation, November 23, 2017, https://theconversation.com/toothless-pan-african-parliament-could-have-meaningful-powers-heres-how-87449.

[33]"Mapping of AU Decision Making Actors and Proesses," Amani Africa Report NO.12, April 28, 2022, https://amaniafrica-et.org/wp-content/uploads/2022/04/Mapping-of-AU-decision-making-actors-and-processes.pdf, 20

[34]Ololade Shyllon and Busingye Kabumba, "Chapter 9: Soft Law and Legitimacy in the African Union: The Case of the Pretoria Principles on Ending Mass Atrocities Pursuant to Article 4(H) of the AU Constitutive Act," essay, in *The Model Law on Access of Information for Africa and Other Regional Instruments: Soft Law and Human Rights in Africa* (Pretoria: Pretoria University Law Press, 2018). Pp 171.

[35]Oliver C. Ruppel and Larissa-Jane Houston, "The Pan-African Parliament of the African Union: Composition, Mandate and Partnerships, and Its Quest for Sustainable Development," African Soil Protection Law, 2021, 485–98, https://doi.org/10.5771/9783748908043-485, 486.

depth of legitimacy doesn't end with participation but also coherence. Coherence is the method in which legitimacy is built by previous documentation, especially previous hard law implementation, to help set a precedent and build legitimacy for future uses.[36] This precedent of legitimacy for expanded powers comes in the form of the African Union directly stating, "The ultimate aim is for the Parliament to be an institution with full legislative powers, whose members are elected by universal suffrage".[37] The PAP's cohesive purpose was to be a legislative body that gained its direct legitimacy from the fact that it was serving all Africans with ideals of Pan-Africanism guiding this continental-wide organization. That purpose, whenever it does come, will lead to the necessary developments to allow the organization to enhance the role of Agenda 2063's soft-law brilliance into implementation by the African people.

Conclusion

Agenda 2063: The Africa We Want, is truly for the Africa and its inhabitants. This was shown through the process of designing the document as "a plethora of stakeholders including civil society groups, women, children, private sector, think tanks, Africans in the diaspora, and the regional economic communities (RECs) were involved and consulted during the development of the Agenda 2063".[38] Thus it would only make sense for an organization with the same value to assist it in its Pan-African implementation. That organization body would of course be the Pan-African Parliament. The institutional deficits that already plague the legislative assembly and get it called "toothless" can be

[36]Ololade Shyllon and Busingye Kabumba, "Chapter 9: Soft Law and Legitimacy in the African Union: The Case of the Pretoria Principles on Ending Mass Atrocities Pursuant to Article 4(H) of the AU Constitutive Act," essay, in *The Model Law on Access of Information for Africa and Other Regional Instruments: Soft Law and Human Rights in Africa* (Pretoria: Pretoria University Law Press, 2018). Pp 171.

[37] "The Pan-African Parliament," The Pan-African Parliament | African Union, February 10, 2022, https://au.int/en/pap.

[38]Eghosa Ekhator, "Sustainable Development and the AU Legal Order," *The Emergent African Union Law*, October 7, 2021, 335–58, https://doi.org/10.1093/oso/9780198862154.003.0019. Pp. 342.

fixed in this bid to also save the enormous potential of Agenda 2063.[39] Without this change, the agenda will just be another documentation that attempts to guide Africa into the light of a more promising tomorrow. Yet, the process of sharpening the teeth of the parliament will take some time. The European Parliament (EP), for which the PAP is quite similar in both goal and supranational coverage, took around twenty-nine years to gain significant powers.[40] This happened through multiple treaties such as the Maastricht Treaty and the Amsterdam Treaty, which mimics the current progress of administrational development which the PAP is going through. That development is through protocols such as the 2014 PAP Protocol which gave the PAP the power to "draft model laws to the Assembly".[41] Thus with similar purposes on similar paths, the EP can act as a roadmap for the PAP in its journey to reach its destiny of an institution with complete legislative powers. As previously mentioned, this will take some time but could also impeded by the actions of African state leaders. While the developmental potential does have the opportunity to help quell continental disputes that could impact sustainable development such as the dispute over the Renaissance Dam in Ethiopia. This would also mean the relegation of power away from heads of state who reside in the AU's Assembly of Heads of State and Government, which currently hold the true power of the AU. The hindrance that follows would delay the agenda further and potentially perpetually. Nonetheless, the enrichment of the PAP is the only way to ensure the survival and continual success of Agenda 2063 while also keeping it in control

[39]Babatunde Fagbayibo, "Toothless Pan-African Parliament Could Have Meaningful Powers. Here's How," The Conversation, November 23, 2017, https://theconversation.com/toothless-pan-african-parliament-could-have-meaningful-powers-heres-how-87449.

[40]Saki Mpanyane, *Transformation of the Pan-African Parliament* (Pretoria: Institute for Security Studies (ISS), 2009). Pp. 6.

[41]Babatunde Fagbayibo, "Toothless Pan-African Parliament Could Have Meaningful Powers. Here's How," The Conversation, November 23, 2017, https://theconversation.com/toothless-pan-african-parliament-could-have-meaningful-powers-heres-how-87449.

of African actors. The legitimacy is present, the backing of the people and the precedent of former parliaments is there. All that is left is for African actors to empower themselves to make Africa the place they want.

Bibliography

"About the African Union." About the African Union | African Union, February 10, 2022. https://au.int/en/overview.

"African Continental Free Trade Area (AfCFTA): Auda." NEPAD. Accessed November 20, 2023.https://www.nepad.org/agenda-2063/flagship-project/african-continental-free-trade-area-afcfta.

"Agenda 2063 - Report of the Commission on the African Union Agenda 2063." African Union, January 31, 2015. https://portal.africa-union.org/DVD/Documents/DOC-AU-WD/Assembly%20AU%205%20(XXIV)%20_E.pdf.

"Agenda 2063: The Africa We Want." Agenda 2063: The Africa We Want. | African Union, January 1, 2019. https://au.int/en/agenda2063/overview.

Amani. "Mapping of AU Decision Making Actors and Processes." Amania Africa Report, April 2022.https://amaniafrica-et.org/wp-content/uploads/2022/04/Mapping-of-AU-decision-making-actors-and-processes.pdf.

Annan, Kofi. "Secretary-General Presents His Annual Report to General Assembly | UN Press." United Nations, September 20, 1999. https://press.un.org/en/1999/19990920.sgsm7136.html.

"AU Rolls out Framework for Monitoring and Evaluating the Status of African Regional Integration." African Union, November 20, 2023. https://au.int/en/articles/au-rolls-out-framework-monitoring-and-evaluating-status-african-regional-integration.

Cofelice, Andrea. "African Continental Free Trade Area: Opportunities and Challenges." *The Federalist Debate* 31, no. 3 (2018): 32–35. https://doi.org/10.2478/tfd-2018-0032.

Ekhator, Eghosa. "Sustainable Development and the AU Legal Order." *The Emergent African Union Law*, October 7, 2021, 335–58. https://doi.org/10.1093/oso/9780198862154.003.0019.

Fagbayibo, Babatunde. "Toothless Pan-African Parliament Could Have Meaningful Powers. Here's How." The Conversation, November 23, 2017. https://theconversation.com/toothless-pan-african-parliament-could-have-meaningful-powers-heres-how-87449.

"Historic New Sustainable Development Agenda Unanimously Adopted by 193 UN Members." United Nations, September 25, 2015. https://www.un.org/sustainabledevelopment/blog/2015/09/historic-new-sustainable-development-agenda-unanimously-adopted-by-193-un-members/.

Hodali, Augustin. THE IMPLEMENTATION OF THE RESPONSIBILITY TO PROTECT (R2P) NORMS BY THE AFRICAN STANDBY FORCE IN SUB-SAHARAN AFRICA, June 9, 2017. https://apps.dtic.mil/sti/pdfs/AD1038733.pdf.

Kaltenborn, Markus. "Implementing International Social Protection Initiatives in Africa – the Role of Global and Regional Soft Law." *Recht in Afrika* 20, no. 1 (2017): 3–10. https://doi.org/10.5771/2363-6270-2017-1-3.

Kioko, Ben. "The Right of Intervention under the African Union's Constitutive Act: From Non-Interference to Non-Intervention." *International Review of the Red Cross* 85, no. 852 (2003): 807–26. https://doi.org/10.1017/s0035336100179948.

Kuwonu, Franck. "Agenda 2063 Is in Harmony with Sdgs | Africa Renewal." United Nations, December 2015. https://www.un.org/africarenewal/magazine/december-2015/agenda-2063-harmony-sdgs.

"Linking Agenda 2063 and the SDGs." Linking Agenda 2063 and the SDGs | African Union, February 10, 2022. https://au.int/agenda2063/sdgs.

Maliszewska, Maryla. "The African Continental Free Trade Area: Economic and Distributional Effects." World Bank Group. Accessed November 20, 2023. https://www.wto.org/english/thewto_e/acc_e/afcfta_feb11maryla.pdf.

"Mapping of AU Decision Making Actors and Processes." Amani Africa Report NO.12, April 28, 2022. b https://amaniafrica-et.org/wp-content/uploads/2022/04/Mapping-of-AU-decision-making-actors-and-processes.pdf.

Mpanyane, Saki. *Transformation of the Pan-African Parliament.* Pretoria: Institute for Security Studies (ISS), 2009.

Otoo, Kwabena Nyarko. "Africa's Economic Trade-Off." Africa's economic trade-off – Economy and ecology | IPS Journal, October 18, 2021. https://www.ips-journal.eu/topics/economy-and-ecology/african-continental-free-trade-area-5496/.

"Our Aspirations for the Africa We Want." Our Aspirations for the Africa We Want | African Union, February 10, 2022. https://au.int/en/agenda2063/aspirations.

"The First-Ten Year Implementation Plan." The First-Ten Year Implementation Plan | African Union, February 10, 2022. https://au.int/en/agenda2063/ftyip.

"The Pan-African Parliament." The Pan-African Parliament | African Union, February 10, 2022. https://au.int/en/pap.

PSC Report. "Staying on Target to Silence the Guns by 2030." ISS Africa, June 22, 2022. https://issafrica.org/pscreport/psc-insights/staying-on-target-to-silence-the-guns-by-2030.

"Regional Economic Communities." United Nations Economic Commission for Africa. Accessed November 20, 2023. https://archive.uneca.org/oria/pages/regional-economic-communities.

Ruppel, Oliver C., and Larissa-Jane Houston. “The Pan-African Parliament of the African Union: Composition, Mandate and Partnerships, and Its Quest for Sustainable Development.” *African Soil Protection Law*, 2021, 485–98. https://doi.org/10.5771/9783748908043-485.

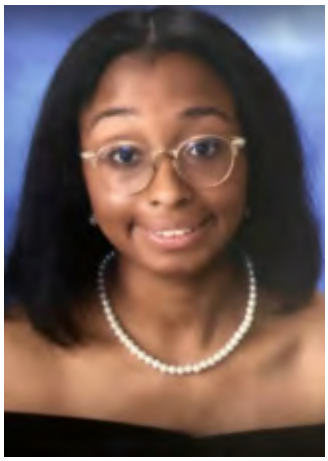
“Second Continental REPORT.” African Union, February 2022. https://au.int/sites/default/files/documents/41480-doc-2nd_Continental_Progress_Report_on_Agenda_2063_English.pdf.

Shyllon, Ololade, and Busingye Kabumba. “Chapter 9: Soft Law and Legitimacy in the African Union: The Case of the Pretoria Principles on Ending Mass Atrocities Pursuant to Article 4(H) of the AU Constitutive Act.” Essay. In *The Model Law on Access of Information for Africa and Other Regional Instruments: Soft Law and Human Rights in Africa*. Pretoria: Pretoria University Law Press, 2018.

Soft law - OECD. Accessed November 20, 2023. https://www.oecd.org/gov/regulatory-policy/irc10.htm.

World Superpowers 2023. Accessed November 20, 2023. https://worldpopulationreview.com/country-rankings/world-superpowers.

Meet the Author:



My name is Arciéne Octavia Bonner, I am a sophomore from Charlotte, North Carolina. I am a UNC-NUS Dual Enrollment student, majoring in Global Studies with a concentration in Global Health & Environment, as well as a world area focus in Asia, studying Chinese. I am a Health & Society and Medical Anthropology double minor and on the pre-med track. Planning on attending medical school, I hope to pursue a specialty in OB/GYN, psychiatry, or anesthesiology. Hoping to do so, when provided the opportunity in my SOCI 469 (Health & Society) course, to conduct research into an aspect of medicine and health, from the perspective of health sociology. I decided to dedicate my research paper to the discussion of Black Maternal Health, having a vested interest in the ever-present, yet continually ignored, health disparities present in the American healthcare system, especially maternal health. As maternal health is an essential aspect of understanding health in the United States, with maternal health intersecting not only reproductive health but pediatric health, as well as women's health. Maternal health in the United States reflects the nation's health disparities, which exist along racial lines, most notably, the Black-White health gap, socioeconomically, as well as regionally, with the American South having worse maternal health outcomes overall.

Black Maternal Health: A Window Into U.S. Health Disparities

Arciéne Octavia Bonner

Abstract

Purpose The purpose of this research paper is to understand factors that contribute to the racial and ethnic disparities present in the United States maternal health, utilizing the White and Black racial binary as it is most reflective of such disparities.

Description In 2018, the United States ranked 32nd out of the 34 Organization for Economic Cooperation and Development nations for maternal health due to its high avoidable maternal mortality rate, which has disproportionately affected Black women. Particularly, Black women of lower socioeconomic status. This paper utilizes multiple studies conducted regarding maternal health, with an array of focuses, such as prenatal care, postpartum care, and motherhood, as such diversity allows for an exploration of maternal health in all respects from beginning to end.

Assessment This paper utilizes mortality and morbidity to determine the causality between race and maternal health outcomes as well as mortality risk to assess the severity of racial disparities. Utilizing the ICD-10 to categorize maternal death or pregnancy-related death.

Conclusion The research paper, from the utilization of the studies included in this effort, found that the United States' White and Black maternal health disparity, results from a multitude of factors; it is this result of socioeconomic factors that determine accessibility to adequate care and arguably most significantly the result of deeply embedded structural and systematic racism, that determines the narratives surrounding Black women's pain and leave them to bear the burden of single motherhood. This results in deleterious health outcomes during maternity, which persists throughout motherhood.

Introduction

The health disparities in the United States that exist between racial and ethnic divides are well documented; with the Black and White divide serving as the primary method of analyzing racial disparities in the country, as race, as a social construct was defined in this binary, defining Whiteness in opposition to Blackness. These health disparities are often compounded by socioeconomic status, SES, which defines one's accessibility to health care resources, from the type of health insurance at one's disposal to the types of surgeries allotted to one's need. Disparities are deeply ingrained into our health infrastructure, with the COVID-19 pandemic, bringing these routinely normalized issues to the forefront of public health discussions. An area of health that can be utilized to analyze such disparities is maternal health, the health of birthing individuals, from prenatal care to postpartum care; maternal health plays an integral role in our society. Maternal health's importance in assessing disparities comes from its intersection with reproduction, reproductive rights, sexism, women's rights, and racial and ethnic inequalities. It is reflective of the many inequalities that disenfranchise marginalized peoples, especially women of color. The United States was ranked 32nd out of the 38 Organization for Economic Cooperation and Development (OECD) nations in 2018, a result of the high avoidable maternal mortality rate disproportionately affecting Black women; and Black women of lower socioeconomic status (SES) in particular. Why does the United States diverge from the directly proportional relationship between income and maternal mortality rate? And, why are Black women in the United States at higher risk of maternal mortality? In this paper, I seek to understand the racial and ethnic disparities present in United States maternal health. The Black and White racial binary, which is most reflective of such disparities will be utilized in conjunction with morbidity and mortality risk, as a method of determining casualty.

Theoretical Model

Health disparities, as defined by the Centers for Disease Control (CDC) are differences that socially disadvantaged populations encounter in the burden of disease, injury, violence, or opportunity to reach optimal health. These differences in health are closely linked to economic, social, and environmental factors that influence one's accessibility to adequate health care. The conceptual model, Figure 1, demonstrates pathways to racial and ethnic disparities in maternal health, specifically severe morbidity and mortality. This combination of social determinants impacts an individual's health status prior to pregnancy, determining their susceptibility to clinical comorbidities and pregnancy complications, which may result in severe maternal morbidity and mortality. The cycle depicted in this conceptual model is representative of the different stages of pregnancy, which include preconception care, antepartum care, intrapartum care, and postpartum care.

1. **Preconception care:** Care provided prior to pregnancy, to identify and modify biological, behavioral, and social risks to a woman's health or pregnancy outcome.
2. **Antepartum care:** Care provided prior to childbirth.
3. **Intrapartum care:** Care provided from the onset of labor to placenta delivery.
4. **Postpartum care:** Care provided after childbirth.

The quality of care provided across this continuum impacts one's risk of severe maternal morbidity and mortality. Race, in health, holds indisputable social significance, as seen with the

historical presence of racial disparities in maternal health. These disparities result in inequalities in healthcare delivery at the system, provider, and patient levels.

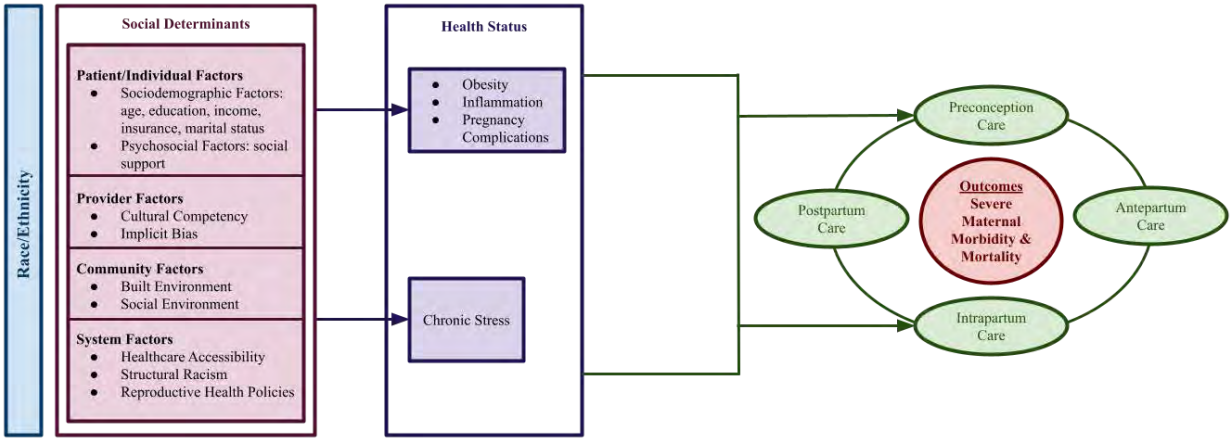


Figure 1. Conceptual Model **(Glazer & Howell, 2021)**

A Brief History: Gynecology A Story of Black Mothers

In order to understand this Black and White gap in maternal health in the United States, it is important to have historical contextualization, as the gap is rooted in history and is a profound example of the perpetuation of historical processes in this country. More specifically it requires an examination of the history of modern obstetrics gynecology (OB/GYN). Though both obstetrics and gynecology, are medical fields that relate to female physiology and medical care and are considered to be one specialty, there are distinct differences between the two. Obstetrics (OB) involves care provided before conception, during pregnancy, childbirth, and immediately following delivery. Gynecology (GYN) involves care of all women's health issues. However, the commonality between the two is their focus on female physiology and their role in maternal health. OB/GYN is the history of American enslavement, it is the history of the medical exploitation of Black women in the United States, as seen by the work of the ‘Father of Modern Gynecology’, a Southern American surgeon, James Marion Sims. Sims is often compared to the likes of the infamous Josef Mengele, a Nazi physician, who similar to many other Nazi party

members, got their inspiration from racist works and practices of the United States. The comparison is derived from his mutilation of enslaved women's bodies in his pursuit of developing new gynecological examinations and surgeries to treat medical issues, such as vesicovaginal fistulas, without the utilization of anesthesia, as he did not see the need for utilizing anesthesia on bodies he did not deem as feeling pain. This notion of Black bodies not experiencing pain to the same extent as White bodies continues to persist to this day, for instance, in a study conducted by Stanton et al., "patients were asked to report how much pain they were experiencing, and physicians were asked to rate how much pain they thought the patients were experiencing", with "physicians [being] more likely to underestimate the pain of Black patients (47%) relative to non-Black patients (33.5%)" (Hoffman et al., 2016).

Pregnancy and Pain: Whose Pain Matters?

Such racially biased disparity in pain interpretation in medical settings is translated into maternal health, in the case of labour pains. In a study conducted by Mathur et al. (2020) on cultural conceptions of labour pain and labour pain management, it was found that even though "White American women were thought to have significantly more labor pain than all women of color", which in the case of this study encompasses African American, Asian, and Hispanic women (Mathur et al. 2020). "African Americans reported greater pain sensitivity compared to both Hispanic and White Americans" (Mathur et al. 2020). This demonstrates the sociocultural factors that impact maternal health outcomes, as it is not physiology that determines the perception of pain. But rather, the implicit biases of those examining pain sensitivity, as seen with the differing perceptions of labour pain by physicians and others, and the individual who is experiencing said pain, when racial biases on pain sensitivity are interpreted as physiological fact. Rather than social constructions built on racism, which in itself is socially constructed, only

holding physiological validity, when its pervasiveness results in deleterious health outcomes for individuals. These disparities in pain management remain in postpartum care, with “Hispanic and non-Hispanic Black women significantly less likely to receive an opioid prescription at discharge compared with non-Hispanic White women” (Badreldin et al. 2019). This demonstrates the pervasiveness of the racial and ethnic disparities present in maternal health in the United States, as these disparities in pain perception and as a result pain management do not end once one gives birth, they persist into one's postpartum care. It is for this reason that it is essential to examine this disparity at multiple stages of maternal health, which includes preconception, antepartum, intrapartum, and postpartum care. These racialized perceptions and management of pregnancy-related pains, such as labor and postpartum pain, relate to important measures of causality between race and maternal health outcomes, those being mortality and morbidity.

Community and Class: Is it Zipcode or Genetic Code?

One’s zip code is a better predictor of one's health than one's genetic code (Graham 2016). One of the most important factors affecting an individual’s life expectancy and overall health outcomes is where they reside. It is argued that one’s zip code alone determines up to 60% of one’s health; this significant role of zip code in overall health outcomes is partly the result of the characteristics in which people live (Graham 2016). Such characteristics include but are not limited to wealth, the level of community investment, and the presence of adequate, well-funded hospitals and health facilities. The social and built environments in which one resides affect the health of one's community and one’s health. An effect that is over and above that of individual behaviors and clinical care. Community health inequities have been made more evident by COVID-19 and are now very seriously out of control. The persistent problem of low- and middle-income (LMI) communities, including rural communities, has been brought to light by

COVID-19. The lack of hospitals in rural and high-poverty areas is one of the main causes of this low investment. Hospitals, even when present in these communities, are typically small and lack the necessary funding to improve the quality of life for residents, including their health. These disparities in health at the community level have become all the more apparent. One such health disparity that was made more apparent was maternal health, with the increased risk of experiencing pregnancy-related complications for pregnant women with COVID-19. Overall, there was an increase in maternal deaths during the pandemic, with COVID-19 in 2020 and 2021 contributing to 25% of maternal deaths compared to 2018 and 2019. Compared to White and Hispanic or Latina women, the maternal death rate for Black or African-American women was disproportionately higher. These women disproportionately resided in LMI communities compared to their White counterparts, and these communities were the most affected by COVID-19, whether the result of severe illness or physiological changes that result from chronic stress exacerbated by the pandemic.

Regarding hospitals, it is important to study the quality of care provided by hospitals located within the community, that is if there is even a hospital located within said community, as the quality of care provided by a hospital is instrumental in determining health outcomes. It has been shown that high Black-serving hospitals, which provided delivery service of approximately 24% of all Black deliveries, and medium Black-serving hospitals, which provided an additional 49.7% of all Black deliveries, are more likely to be teaching hospitals, to be in the South, to be in an urban region, to have more deliveries per year, to have bigger beds, and to deliver more Medicaid patients (Howell et al. 2013). High-Black serving hospitals (29.4) and medium-Black hospitals (19.4) have higher severe morbidity rates, than low-Black serving hospitals (29.4 and 19.4 vs 12.2 per 1000 deliveries, respectively; $p<.001$) (Howell et al. 2013). It is important to

recognize that it is not the fact that these hospitals primarily serve a Black population, but rather the lack thereof funding, which affects their ability to provide adequate healthcare services, that results in such disparaging health outcomes. White women who delivered at high Black-serving hospitals experienced an elevated adjusted rate of severe maternal morbidity (19.2 per 1000 deliveries), not experienced at low Black-serving hospitals (Howell et al. 2013). This adjusted rate is similar to that of which is experienced by Black women who delivered at said hospitals (20.5 per 1000 deliveries) (Howell et al. 2013). Such findings demonstrate the necessity of interventions at hospitals located in predominately Black and low- and middle-income communities, as the incorporation of interventions, such as thrombotic therapy, carotid imaging, angioplasty, and provision of timely antibiotics for pneumonia, can have a considerable impact on morbidity and mortality risk.

Politics and Race: Reproductive Rights, Race, and Maternal Health

Intersectionality is a term and method of study, that was developed out of Black feminist thought and theory and popularized by Kimberlé Crenshaw, which asserts that in studying marginalization and discrimination it is necessary to take into consideration everything and anything that can marginalize a people—including but not limited to sex, race, class, sexual orientation, and physical ability. It is for this reason that when studying the Black and White maternal health disparity, it is conducted through an intersectional lens that takes into consideration, racism, and sexism, as such, this paper incorporates the study of the effect of reproductive rights policies on American women, specifically how it affects Black women in comparison to their White female counterparts. Accessible family planning services, prenatal care, and safe abortion access are listed by international organizations as the three main ways to lower maternal mortality. These modes of lowering maternal mortality overlap as the presence of

one necessitates the presence of the other, as seen with there being a high maternal morbidity and mortality rate in states that have restrictive laws that make abortions within the health system inaccessible and increase unsafe abortions outside the health system. The presence of restrictive reproductive laws represents the presence of negative attitudes toward women and as a result their reproductive health. Such restrictive reproductive laws result in negative maternal health outcomes, with states that enacted restrictive abortion legislation experiencing a 38% increase in maternal mortality rate. The closure of Planned Parenthood clinics contributed to the overall increase in maternal mortality in the United States, an increase that disproportionately affects Black women. As 35% of patients at Planned Parenthood health centers are Black or Latinx, and 56% are in medically underserved or rural areas; 75% have incomes at or below 150% of the federal poverty level (Patterson et al. 2022).

Additionally, 60% of Planned Parenthood patients receive care through Medicaid or the Title X family planning program, which gives low-income people access to contraception and other reproductive healthcare (Patterson et al. 2022). In a study conducted by Patterson et al. (2022), it was found that Black women in states with supportive reproductive rights laws, though still faring worse than their White counterparts, had a significant improvement in their maternal mortality risk. Demonstrating that policy changes that are antagonistic to women's health, disproportionately affect Black women, as they are as a result of intersectionality, racism, and sexism, the most vulnerable to the effects of said policies.

Morbidity and Mortality: Correlation or Casualty?

Morbidity and mortality, while similar and frequently associated, are not the same. Morbidity refers to the state of being symptomatic or sick with a disease or condition. Mortality refers to the number of deaths brought on by the health incident under investigation, which in this case would be pregnancy. Morbidity is typically represented by prevalence or incidence, while a rate per 1000 people or an absolute number typically represents mortality. In short, morbidity refers to having a disease or condition, while mortality refers to death resulting from a disease or condition. Maternal mortality in this paper utilizes the International Statistical Classification of Diseases and Related Health Problems to determine maternal and pregnancy-related deaths, utilizing ICD-10 (1995 to 1998) and ICD-9 (1999 to 2013). The ICD-9 categorizes deaths with codes 630 to 676 as maternal deaths. While the ICD-10 categorizes maternal deaths with codes A34, O00 to O-95, O98 to O99, and P00 to P96. These codes determine maternal and pregnancy-related deaths and allow for the investigation of maternal mortality and mortality risks.

Regarding both morbidity and mortality, numerous studies demonstrate significant racial disparities in medical morbidity and mortality in the U.S., with non-Latinx/Hispanic Black women performing significantly worse than their non-Latinx/Hispanic White counterparts (Patterson et al. 2022). Black women are at greater risk of morbidity and mortality than White women. This disparity is translated into maternal health, as seen with, even when controlling for age and women's reproductive rights support, the maternal mortality risk of Black women is typically doubled that of White women (Patterson et al. 2022). The maternal mortality rates of Black women in their early twenties are comparable to those of White women in their mid-to late-thirties or older. This is concerning, considering that female mortality rates and risks tend to be lower than their male counterparts (Patterson et al. 2022). However, even when not solely

focusing on maternal mortality risks and rates, on average Black women are at higher risk of mortality than their Black male counterparts, and their White female and male counterparts. This holds for morbidity risk as well, which could be argued to be the result of women having a higher risk of morbidity on average. However, in terms of maternal health, this argument does not hold. As such, this occurrence of Black women experiencing an increased risk of morbidity and mortality than their White counterparts results from a mixture of racism and sexism.

The Longterm Effect: Telomere Length

Though this research paper focuses on the Black and White disparities in maternal health, through the analysis of socioeconomic, sociopolitical, and sociocultural factors, those being patient, provider, community, and system factors, such factors have disproportionate effects on health outcomes along racial lines. Specifically their morbidity and mortality risk, it is necessary to determine if such deleterious health outcomes continue after their postpartum period. Telomere length has been utilized for this purpose, as it is a sound biomarker of aging, allowing it to represent the detriment of maternal health disparities through stress on one's life expectancy. This paper utilizes the results attained by a study conducted by Niño et al, in 2022, which sought to answer the following questions:

1. "Are there divergent racial and ethnic patterns of cellular aging among mothers that have experienced paternal incarceration" (Niño et al. 2022)?
2. "How do secondary stressors, such as economic hardship, neighborhood concentrated poverty, maternal mental health, and parenting stress shape paternal incarceration-cellular aging patterns for Black, Latin[x], and White mothers" (Niño et al. 2022)?

Both studies found that Black mothers were more likely than their Latinx and White counterparts to report that their male partner, the father of their child[ren] was incarcerated, and they were

also found to be exposed to the three indicators of economic instability namely material hardship, poverty, and neighborhood concentrated poverty. Paternal incarceration, as a stressor in itself, was not found to have any significant association with the telomere length of Black, Latinx, or White mothers. However, data showed that exposure to paternal incarceration had statistically significant and adverse relationships with cellular aging after controlling for multivariate factors that were probably related to paternal incarceration and maternal health, but only for Black women (Niño et al. 2022). This demonstrates that exposure to paternal incarceration accelerated cellular aging among Black mothers ($p<.05$), which is not present in Latinx and White mother populations.

This research demonstrates the effect of systemic factors, that being the mass incarceration of Black men, patient factors, those being race and economic status, and community factors, that being neighborhoods with concentrated poverty, on maternal health outcomes that continue from pregnancy into motherhood. The study found that race and ethnic background determine the risk of detrimental physiological consequences of paternal incarceration on mothers. This study helped answer my research questions, as it provided a more in-depth analysis of the racial and ethnic disparities in maternal health beyond that of maternal mortality, providing insight into the racialization of our prison industrial complex and how the paternal incarceration of Black men affects the maternal health of Black women at a disproportionate rate than their White counterparts. It also demonstrates the social nature of maternal health and how race and ethnicity can be a determinant of one's health status.

Conclusion

This research paper found that the United State's Black and White maternal health disparity results from a multitude of factors, resulting from the patient, provider, community, and system factors that determine accessibility to adequate care. Most significantly, the result of deeply embedded structural and systematic racism, determining narratives surrounding Black women's pain and leaving them to bear the stressful burden of single motherhood. It is racism and sexism that result in deleterious health outcomes before pregnancy, compounded by pregnancy, and persist after pregnancy into motherhood. As such, research on maternal health must incorporate race as a social variable to provide an in-depth investigation and comprehension of the disproportionate distribution of these risks. This study is limited in brevity, as it is not a longitudinal one, studying maternal health beyond pregnancy, would allow for a greater understanding of the pervasiveness of the Black and White maternal disparity and its effects on health outcomes, such as chronic stress from motherhood and inflammatory disease, thus morbidity and mortality risks. Such a study would utilize telomere length to better understand the effect of chronic stress on the aging process, resulting from socioeconomic factors that develop health disparities that disproportionately affect Black mothers.

References

Badreldin, N., Grobman, W. A., & Yee, L. M. (2019). Racial disparities in postpartum pain management. ***Obstetrics & Gynecology***, ***134***(6), 1147–1153. https://doi.org/10.1097/aog.0000000000003561

Debbink, M. P., Ugwu, L. G., Grobman, W. A., Reddy, U. M., Tita, A., El-Sayed, Y. Y., Wapner, R. J., Rouse, D. J., Saade, G. R., Thorp, J. M., Jr, Chauhan, S. P., Costantine, M. M., Chien, E. K., Casey, B. M., Srinivas, S. K., Swamy, G. K., Simhan, H. N., & Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) Maternal-Fetal Medicine Units (MFMU) Network (2022). Racial and Ethnic Inequities in Cesarean Birth and Maternal Morbidity in a Low-Risk, Nulliparous Cohort. Obstetrics and Gynecology, 139(1), 73–82. https://doi.org/10.1097/AOG.0000000000004620

Glazer, K. B., & Howell, E. A. (2021). A way forward in the maternal mortality crisis: Addressing maternal health disparities and Mental Health. ***Archives of Women's Mental Health***, ***24***(5), 823–830. https://doi.org/10.1007/s00737-021-01161-0

Graham G. N. (2016). Why Your ZIP Code Matters More Than Your Genetic Code: Promoting Healthy Outcomes from Mother to Child. ***Breastfeeding Medicine: the official journal of the Academy of Breastfeeding Medicine***, 11, 396–397. https://doi.org/10.1089/bfm.2016.0113

Hardeman, R. R., Kheyfets, A., Mantha, A. B., Cornell, A., Crear-Perry, J., Graves, C., Grobman, W., James-Conterelli, S., Jones, C., Lipscomb, B., Ortique, C., Stuebe, A., Welsh, K., & Howell, E. A. (2022). Developing tools to report racism in maternal health for the CDC Maternal Mortality Review Information Application (MMRIA): Findings

from the MMRIA Racism & Discrimination Working Group. ***Maternal and Child Health Journal***, ***26***(4), 661–669. https://doi.org/10.1007/s10995-021-03284-3

Hoffman, K. M., Trawalter, S., Axt, J. R., & Oliver, M. N. (2016). Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites. ***Proceedings of the National Academy of Sciences***, ***113***(16), 4296–4301. https://doi.org/10.1073/pnas.1516047113

Howell, E. A., Egorova, N., Balbierz, A., Zeitlin, J., & Hebert, P. L. (2016). Black-white differences in severe maternal morbidity and site of care. ***American Journal of Obstetrics and Gynecology***, ***214***(1). https://doi.org/10.1016/j.ajog.2015.08.019

MacDorman, M. F., Thoma, M., Declcerq, E., & Howell, E. A. (2021). Racial and ethnic disparities in maternal mortality in the United States using Enhanced Vital Records, 2016–2017. ***American Journal of Public Health***, ***111***(9), 1673–1681. https://doi.org/10.2105/ajph.2021.306375

Mathur, V. A., Morris, T., & McNamara, K. (2020). Cultural conceptions of women's labor pain and Labor Pain Management: A mixed-method analysis. ***Social Science & Medicine***, ***261***, 113240. https://doi.org/10.1016/j.socscimed.2020.113240

Mi, T., Hung, P., Li, X., McGregor, A., He, J., & Zhou, J. (2022). Racial and ethnic disparities in postpartum care in the greater Boston area during the COVID-19 pandemic. JAMA Network Open, 5(6). https://doi.org/10.1001/jamanetworkopen.2022.16355

Nelson, D. B., Moniz, M. H., & Davis, M. M. (2018). Population-level factors associated with maternal mortality in the United States, 1997–2012. ***BMC Public Health***, ***18***(1). https://doi.org/10.1186/s12889-018-5935-2

Niño, M., Harris, C. T., Tsuchiya, K., & Hearne, B. (2022). Paternal incarceration, Race and

ethnicity, and maternal health. ***Journal of Racial and Ethnic Health Disparities***. https://doi.org/10.1007/s40615-022-01388-2

Patterson, E. J., Becker, A., & Baluran, D. A. (2022). Gendered racism on the body: An intersectional approach to maternal mortality in the United States. ***Population Research and Policy Review***, ***41***(3), 1261–1294. https://doi.org/10.1007/s11113-021-09691-2

Meet the Author:

Erin Mazur is a graduating senior from Kennett Square, PA at the University of North Carolina at Chapel Hill, and she is currently pursuing her Bachelor of Science in Biology. In the Fall, she will be entering the Eshelman School of Pharmacy to pursue a PharmD. Erin has been heavily involved in research through a research assistant position in Dr. Prem Lakshmanane's Immunology Lab, a Microbiology research internship, and more. Erin's JOURney submission is the culmination of her work during her UNC SLATE research fellowship. This fellowship allowed her to pursue an independent project of her choosing, relating to racial equity of Africans. She studied the intertwinings of neocolonialism, public health research, and evidence-based medicine in sub-Saharan Africa. At large, her research interests include vaccine development, personalized medicine, and the socioeconomic determinants of health.

Title: The discovery-care continuum: an avenue to dismantle neocolonialism in global health

Abstract

The global health field tackles health inequities and, therefore, requires inequities to be present for the field to continue to exist. Thus, despite the field's tremendous work, it turned to exploitative research to maintain a level of inquiry and preserve its existence. The emergence of evidence-based medicine (EBM) significantly worsened the exploitative research practices of global health, causing researchers to value data over the establishment of proper health care in certain areas. This mass data collection did not improve health care because the findings were poorly translated into practice, even if research focused on diseases prevalent in low and middle-income countries (LIMC). Solving this research-clinical practice gap is of high importance in low-resource areas in Sub-Saharan Africa (SSA), where high disease burden and economic and social conditions significantly affect health. Academic health science centers offer a way to combine the elements of the discovery-care continuum–research, education, and practice to solve this research-clinical practice gap. If these centers are organized into international systems, they are well-positioned to address health disparities in LIMC; however, current research is lacking to support this capacity. Confronting this lack of evidence, this project analyzes five public health school research-practice frameworks. The project deduces that sharing faculty and more granular projects increase the success of integrating the discovery-care continuum.

Key Words

Global health, Neocolonialism, Research, Evidence-based Medicine, Sub-Saharan Africa

Introduction

The field of global health was established to reduce health disparities, but to maintain the field, research practices must maintain said inequities, making it a neocolonialist venture[1]. These issues were further exacerbated by evidence-based medicine (EBM), which popularized exploiting low-resource countries as research settings for randomized control trials (RCT)[2]. Even if ethically sourced, evidence for EBM is poorly translated into care, a concern termed the "research-clinical practice gap"[3]. While the current practices of EBM are not optimal, its mission to use the best available evidence to improve health outcomes is important[4,5]. To improve the EBM system, public health initiatives must find a consistent way to close the research-clinical practice gap, as this would ensure data use instead of just data collection[6]. Closing this gap is especially important in Sub-Saharan Africa (SSA), a region burdened by co-infection, poverty, and conflict, and therefore requiring differentiated service, not one-size-fits-all medicine[7,8].

In the US, academic health science centers (AHSCs) became prominent shortly after the advent of EBM. With a tripartite mission to research, educate, and provide quality care, AHSCs are uniquely positioned to ensure that research findings complete the "discovery-care continuum," a linear progression from proof of concept to implementation and policy[9,10]. Their bidirectional structure allows for research concentration on the specific challenges clinicians face[11]. Reorganization into international systems would enable academic health science systems (AHSSs) to close the research-clinical practice gap and positively impact health inequities by integrating Western research funds with the provision of care in LIMC[10].

Building these systems is no small feat, requiring careful alignment of research priorities, community engagement, and funding[12].

Public health schools, as a consortium of researchers, educators, and humanitarian workers focused on eliminating health inequity, provide a

natural framework for closing the research-clinical gap in SSA through AHSSs. However, evidence demonstrating AHSS's contribution to health equity is scarce. Considering this lack of evidence, this project presents five case studies of different research-clinical practice frameworks from top U.S. global health programs. The study analyzes their work in meeting targets within the United Nations' Sustainable Development Goal 3: good health and well-being for all[13]. Framework comparison produces guidelines for successful research translation in SSA low-resource settings.

Colonialism, the Advent of Global Health, and Evidence-Based Medicine

Under the guise of a "developing" and "civilizing" mission, European colonialist practices forced the production of cash crops and depleted African colonies of raw materials, arresting their natural economic development and leaving them resource-poor[14]. This legacy of economic inequity affected health care, with previously colonized countries suffering from higher disease rates, reduced capacity to address epidemic diseases, and inability to afford treatments[15,16].

Global health emerged as a field in 2003 to sponsor human immunodeficiency virus (HIV) treatment in Africa, funding free antiretroviral treatment (ART) in low-income countries[17]. The President's Emergency Plan for AIDS Relief (PEPFAR) approach did not scale up HIV services or address the more significant needs of the health system and instead, created separate health systems dedicated only to HIV. This intervention perpetuated aid dependency rather than building African LIMC's healthcare capacity[18]. This preservation of inequality persisted in global health research practices, which began to use previously colonized settings as research platforms. African LIMC provides a resource-poor, drug-naive, and high disease-burden area[15]. These conditions offer Western researchers continuous funding and publication opportunities, diverting money back to the West. This cycle would not be secure if local

health and research facilities were stronger[17]. While individual researchers may be motivated to improve health in LIMC, global health as a system is more motivated to minimize epidemics' impact on the Global North and is latently tasked with maintaining racial hierarchies.

These practices are exacerbated with the use of EBM. Because EBM only values statistical and experimental models of evidence produced through RCT, its predominance has direly affected global health, which entered an era of providing experimental medicine or none at all[2]. For example, comparing the success of a marketing strategy to sell condoms versus the distribution of free condoms for STD prevention is considered more useful under EBM than documenting lower STD rates in areas where free condoms were distributed[2]. Additionally, the “magic bullet” approach of EBM puts forth that medical technological advances could overcome the social conditions that affect the quality of healthcare, perpetuating the notion that it is ethical to use LIMCs for research without capacity building[2].

The Research-Clinical Practice Gap

Clinicians widely accept EBM because, in principle, it improves health outcomes by utilizing the highest quality of evidence, increasing care consistency, and reducing costs[4,19]. For example, proponents cite that care outcomes improve by 28 percent when based on the best evidence[20].

However, EBM manifests quite differently, producing a tension between the conduction of research and care delivery termed the research-clinical practice gap[21]. EBM promotes uniformity of care and fails to account for the complexities of health care[22]. Population-based research needs to be more applicable to the individual patient, and EBM compels clinicians to use the most effective treatment, according to the statistics, instead of integrating knowledge of the patient and available evidence to decide what is best[22,23]. Even in the best health systems, EBM is difficult to implement due to the vast results that database searches provide. Therefore, it is

hard to identify the best evidence, and clinicians strained by high caseloads need more time to critically analyze the available evidence [24]. Additionally, research from EBM often lacks guidelines for the translation of results to clinical practice[25].

The current research practices associated with EBM undermine the value of the evidence generated. There is a strong cultural bias toward Western research priorities[21]. For example, only 10% of global health research funds are spent to study and fight the diseases that cause 90% of the world's health burden[4]. With South Asian and Sub-Saharan African countries leading the world in disease burden, there is a clear prejudicial focus on ailments that impact the West, as opposed to developing research to help the most afflicted countries. There are also frequent instances of unnecessary research, where an RCT is conducted when sufficient evidence already exists[23,26].

The Importance of Closing the Research-Clinical Practice Gap in LIMC and Global Health's Role

The gap between conducting research and improving care is pressing in African LIMC settings. Individualized care, which current EBM practices do not promote, is of utmost importance in African LIMC settings where malnutrition, co-infections, traditional remedies, substandard and counterfeit medicines, and late entrance to care all impact the success of interventions[7,8]. Evidence for the Global North does not apply to these settings because of differences in resources and common ailments[21]. For example, in African LIMC, diseases such as hypercholesterolemia, chronic obstructive pulmonary disease, and depression are undertreated[27-29]. While in the Global North, these diseases are routinely screened for and often respond well to intervention, these practices are not easily implemented in the poorly resourced health systems of SSA[30,31]. EBM is complicated to implement in low-resource settings where overburdened clinicians trying to provide care to large numbers of patients are needed to lead research efforts[21].

The COVID-19 pandemic further

disrupted the feasibility of EBM in African LIMC. For many LIMC residents, community health workers (CHW) are the most accessible form of health care[22]. Fear grew that healthcare workers would spread COVID, and CHW lacked the proper personal protective equipment to continue care. Reduction in care in high disease burden areas along with a novel disease has unprecedented effects on health, producing an urgency to gather and evaluate real-time findings to provide proper care post-pandemic[32]. The urgency of the pandemic has illuminated the deficiencies of traditional research models, such as RCT. It is essential to improve research capacity and health data collection systems in SSA to better adapt to quickly changing conditions in health.

As global health shifts focus towards achieving Sustainable Development Goal 3: Better health and well-being by 2030, using targeted interventions rather than one-size-fits-all medicine in African LIMC is vital[36]. The objective of EBM is to combine reliable science with compassionate care, which would be extremely useful in African LIMC, yet the current system has yet to reach its full potential[37]. Global health must tackle fundamental issues of colonialism's legacy by building research capacity in LIMC and conducting research with local African regions in mind. Developing medical databases would prompt non-experimental research, therefore saving resources for care. Global health projects should fill research gaps; Western organizations can dive into topics that otherwise would not get proper funding from African governments because of other pressing health care needs[38]. These projects should function through partnerships with researchers in LIMC, study prevalent conditions of local concern, and strictly follow an ethical standard[36-38]. Research projects should have clear translation guidelines to ensure the results will be successfully implemented to care[6].

Academic Health Science Systems and the Research-Clinical Practice Gap

In the early 2000s, Academic Health

Science Centers (ASHCs) became prominent in the United States' healthcare system. Usually consisting of a school of medicine, teaching hospital, and tertiary medical centers, these organizations pursue a tripartite mission to conduct research, educate future professionals, and deliver quality patient care[39]. Because they participate in each sector of the discovery-care continuum, they are uniquely positioned to lessen gaps between research findings and translation to patient care[40]. Their bidirectional structure and joint governance allow them to structure research goals based on feedback from their clinicians[41,42]. If ASHCs were reorganized into larger systems of global partners, all would benefit from shared informatics, and their capabilities for translational research, education, and capacity building would increase[43]. Individual members' autonomy would drive innovation through continuous collaboration and competition, instead of stifling it[40]. While many ASHCs focus on their direct community, ASHSs, if scaled up, could integrate with global health organizations and formulate global solutions[10]. With their triple mission and ability to cross-cut university programs, ASHSs are situated to impact the social determinants of health (SDOH), aid low-resource settings, and address health disparities[9,44]. However, there is little evidence in the literature demonstrating a contribution[37].

Confronting this lack of evidence, this project analyzes research-clinical practice frameworks of top global health programs. Johns Hopkins' Bloomberg School of Public Health (BSPH), the University of North Carolina at Chapel Hill (UNC-Chapel Hill) Gillings School of Public Health and Columbia University Mailman School of Public Health (MSPH) were selected based on similar research topics routed in LIMCs and cross-cutting organizations.

Integrating all aspects of the discovery-care continuum through mutual faculty

The minimal progress towards Sustainable Development Goal 3 illustrates

the current shortcomings in translating research through the discovery-care continuum. One goal is to reduce the global maternal mortality average by at least 157 deaths per 100,000 live births[13]. Statistics show that between 2016 and 2020, the global average reduction rate was only 0.04%. With 70% of these deaths occurring in SSA, innovative interventions in these nations are crucial to meeting this goal[45]. SSA must also be targeted to end preventable death under age 5. LIMC only accounts for 50% of the global population under age five but suffers 80% of the deaths[46]. Infections and low birth weight are leading causes of newborn mortality, which are preventable with early initiation of breastfeeding[47,48]. Only three-fifths of infants globally are breastfed in the first hour of life, despite statistics showing that the chance of death increases by 33% by waiting until after the infant's first 24 hours to breastfeed[48]. This data shows there must be more intervention in LIMC to increase early breastfeeding and solve preventable deaths under 5.

To achieve equitable care, better care than current practices is needed. To address these disparities, AHSSs must fulfill the discovery-care continuum to ensure innovation and fast implementation. They must address research, practice guidelines, policy creation, and service delivery, and through this network, evaluate the effectiveness of the delivery and conduct follow-up research. By integrating these steps themselves, AHSSs can reduce the research-clinical practice gap.

UNC Gillings and Johns Hopkins BSPH maternal and newborn health strategies reside under one institution that forms partnerships with various organizations focused on different aspects of the discovery-care continuum. This structure seems promising as it allows each partner to utilize the university's research for their specific purpose and presents a pathway for challenges found in care to be studied. While functional, these schools' current organizations are not reaching the full potential of quickly translating research into practice.

UNC Gillings

One tenet of UNC's Humanitarian Health Initiative is to support best practices in infant and young child feeding to reduce young mortality[49]. At UNC, the Humanitarian Health Initiative (HHI) and Carolina Global Breastfeeding Initiative (CGBI) champion this mission via research and support outside organizations: the Global Breastfeeding Collective and Core Group. The Global Breastfeeding Collective garners political and financial support for breastfeeding and disseminates resources on feeding counseling[50]. Core Group synthesizes current research to provide technical updates to partners and creates guidance and resource materials on topics such as pregnancy care best practices and indicators and maternal nutrition[51]. Organizing and disseminating the best evidence is critical in transitioning from university research to practice. It is also essential that the Collective and Core Group focus on different topics in their resources to cover more ground, but their work does not assure timely implementation. CGBI does hands-on training only in the U.S., so they see the need for this service but do not provide it in SSA, where it is needed most[52]. Additionally, the resources need to reflect current research. Core's most recent resources discussing child mortality and nutrition are from 2013 and 2015, respectively[51]. The Collective's latest counseling course from 2021 cites sources ranging from 2004 to 2016[53]. The need for new research implementation is crucial as the current practices do not have nations on target to meet SDG goals[47]. These organizations' connection to UNC Gillings provides an opportunity to do so, but CGBI research is not specific to LIMC, which, for targeted EBM interventions, is not sufficient[54]. HHI research could be more helpful; they completed maternal health research on breastfeeding in COVID. However, the research was completed after many of the Collective's resources on IYCF in COVID were posted, and the Collective does not cite HHI research[50,55].

There should be a direct association between assessing challenges and the

formation of research. Suppose UNC-Gillings had an avenue of direct service in SSA. In that case, they can better determine what research still needs to be done to solve current issues and, therefore, improve breastfeeding practices to reduce infant mortality on the road to SDG 3.2.

Johns Hopkins Bloomberg

Johns Hopkins Center for Humanitarian Health (CHH) and their university-founded corporation, Jhpiego, focus on maternal and newborn health. While CHH conducts some independent research, most is performed by Jhpiego, which is also responsible for Performance Monitoring for Action (PMA), Momentum, and the Advance Family Planning Initiative (AFI). The three auxiliaries, along with BSPH, cover the discovery-care continuum. PMA collects data on maternal and newborn health indicators[56]. AFI aids in policy enactment, such as fund allocation[57]. Momentum provides direct services such as expanding the range of contraceptives available and ensuring access to obstetric surgeries[58].

There is a clear direction in PMA's research focusing on abortion safety and post-abortion care. PMA measured the general abortion safety in SSA and then researched specific demographics more susceptible to undergoing unsafe abortions[59-61]. Next, PMA assessed the availability and safety of post-abortion care, specifically in areas with a high burden of unsafe abortions[62,63]. This research cascade is well aligned with Momentum's work to increase the quality of post-abortion care in some regions. However, PMA studied abortion care in Nigeria, Côte d'Ivoire, Uganda, Kenya, and the Democratic Republic of Congo (DRC). Momentum works in these countries but only focuses on abortion care in the DRC[64]. It also focuses on abortion care in Mali, but this nation was not included in the study[65]. Better coordination between the regions of service provision and research priorities would allow for more effective change.

Research priorities are better aligned between research and service

delivery under Momentum's theme of ensuring maternal and newborn care in fragile settings. Paul Spiegel is completing his research at BSPH, but he is also leading faculty on a project involved with Momentum[66]. His research studies maternal and child health, specifically in armed conflict areas and supports Momentum's work in South Sudan, which is focused on strengthening the health systems in the newly independent country[67,68]. He also studies the delivery of nutrition interventions in conflict settings, specifically in the DRC, Mali, and South Sudan[69]. The combination of these projects benefits the Mali Momentum program, which tailors nutrition care delivery for the fragile setting[58,65]. His direct engagement in research and projects is associated with better alignment. In addition, he can study the implemented interventions to ensure they are successful, such as in his project that develops indicators on public health interventions in humanitarian crises[68]. Evaluating interventions is essential as it builds the evidence base for best strategies via trial and error.

When care should be context-specific: specificity and follow-through in research

10.3 million people in Sub-Saharan Africa (SSA) are currently untreated for HIV, and there are a projected 1.2 million new cases each year[70]. HIV is an incredibly complex public health issue. The mass stigma against people living with HIV dissuades many from testing and care, as do cost, lack of time, and distrust in the health system[71-73]. In addition to increased supply, testing, and treatment in these low-resource health systems, solving the epidemic will require research and scaling up successful interventions[70,74]. The high incidence of stand-alone HIV clinics from PEPFAR and ART scale-up will not be enough to resolve the complex epidemic[74,75]. Quality HIV care has become context-specific, with individuals requiring maternal and newborn care, tuberculosis care, or mental health care for good outcomes[76]. While EBM suggests that repeated RCTs are the best method for tackling the epidemic as they lead to solid

conclusions, it is unethical to randomize people living with HIV to suboptimal care for the sake of a study. There is a movement to use observational data and leverage data from medical records to complete HIV research as opposed to disrupting care with RCT[75]. While the mass amounts of information this form of research generates are vital for basing clinical decisions on real-time evidence, it still stops efforts at the study and does not continue to ensure education or care provision.

The International Epidemiologic Database to Evaluate AIDS (IeDEA) is a global research consortium with top public health programs as participants, such as Johns Hopkins, UNC Gillings, and Columbia University. The consortium pools data on critical variables in the hopes of simplifying HIV/AIDS research and answering unique questions individual studies cannot address[77]. IeDEA data would greatly impact if translated into clinical service by these academic health science systems; however, data application is inconsistent.

Johns Hopkins Bloomberg

In addition to its maternal and childcare components, Johns Hopkins' Jhpiego funds and diverts resources to Reaching Impact, Saturation, and Epidemic Control (Rise), an HIV and COVID-19 service delivery organization. While Rise is active regarding HIV in fourteen countries, it partners with Johns Hopkins BSPH and the International Center for AIDS Care and Treatment Programs (ICAP) at Columbia University, specifically in Tanzania[78]. ICAP currently studies critical populations for HIV care in Tanzania[79]. BSPH professor Abdullah Baqui is an expert on maternal and newborn health and CHWs in Tanzania[80-82]. He currently leads two projects devoted to integrating HIV care with maternal and newborn care and educating CHW care to ensure women and children's access to HIV care[83]. The combined research of ICAP and Baqui is valuable input for Rise's maternal and child health work in Tanzania on targeting key populations, achieving self-sufficiency of human resources for health, and building the financial capabilities for

ensuring continuous care, as integration of care is a highly cited way to do so[76,78].

The benefits of research efforts are not always translated into clinical practice for targeted populations. IeDEA's study on global site level comprehensiveness and its association with care retention was only cited by two publications, neither belonging to the IeDEA consortium[84]. These researchers did not use the data to continue its progress through the discovery-care continuum. One believes care integration could be effective, but more highly suggests using less dose, longer acting treatments to reduce loss-to-follow-up care, and the other concludes that cervical cancer screening, to be effective, cannot rely on just HIV clinics[85,86]. This impediment to research translation is seen again in an IeDEA publication on HIV and tuberculosis care integration[87]. While cited three times, none of these publications are intervention-directed and instead use the methodology for different research targets[88-90].

Globally relevant research is not as easily translated throughout the discovery-care continuum. Rise's specific project allowed applicable research to be connected more easily. This project exemplifies how it is more rewarding to focus research and care efforts on a specific context and see it through than try to make mass change at once.

UNC Gillings

Even when projects are niche enough to be easily translated into practice, the researchers' decisions on who to partner with can impede implementation. Dr. Angela Parcesepe does distinct research on the interconnectedness of mental health diseases and HIV in Cameroon[91-93]. Her research is funded by IeDEA and is grounded in data collection rather than translation into practice. The publications have no guidelines for disseminating information to clinicians or policymakers in Cameroon. While her research is valuable, it's likely to get caught in the research-clinical practice gap without efforts tied to immediate translation.

On the other hand, Dr. Frieda Behets

researches mother-to-child transmission of HIV in the DRC. While she has one publication within IeDEA, the majority of her work takes place through PEPFAR funding, where she has employed conditional cash transfers to retain women in care to prevent mother-child transmission and where she trains HIV-positive mothers to counsel HIV-positive pregnant women on best practices to reduce transmission which has also been highly successful in keeping the pregnant women engaged in care[94,95]. Despite being only slightly involved in IeDEA, Dr. Behets's work shows that choosing to operate under different funding can open the door to translating one's research into practice and making an immediate impact.

Additionally, UNC's Measure Evaluation organization shows that mass data collection can be productive if attached to a program. PEPFAR launched the DREAMS Initiative in ten SSA countries to reduce HIV incidence in girls and young women, a group that comprises 71% of new HIV infections[96]. Measure Evaluation collected the data to inform the programming, such as identifying high-risk groups' characteristics, developing success indicators, and identifying effective interventions. The data and analysis result from years of data collection from health information systems in ten countries. Mass data is complex to analyze, but with a goal in mind, the triangulated data was successfully organized and applied to an intervention[97]. The collection of mass observational data is not at fault; however, it is necessary to identify a key population and use the data to find solutions for specific groups, which is most easily done when connected to a program already working towards a solution.

Columbia University

While the International Center for AIDS Care and Treatment Programs (ICAP) has expanded to tackle many health challenges, its work in HIV/AIDS is especially noteworthy[98]. Their Accelerating Children's HIV/AIDS Treatment (ACT) Initiative provides non-monetary rewards to adolescents in Kenya for treatment adherence and care retention[99]. They also fostered the

first multi-country HIV treatment program explicitly aimed at low-resource settings[100]. What began as a program to fund treatment for families to stop transmission became a capacity-building organization that now implements population surveys and provides technical support for electronic medical record keeping, allowing LIMC in SSA to be self-sufficient in evidence-based medical decisions[101,102].

Now focusing on capacity building, it is essential that ICAP support innovative research to fuel progress. They successfully do so by leading the HIV Coverage, Quality, and Impact Network (Cquin)[103]. This network convenes global leaders to work together to scale up differentiated service delivery. For example, in Kenya, members are piloting innovative approaches to TB/HIV care integration, and in Ethiopia, they are evaluating the success of adherence clubs for TB/HIV care integration[104]. The individual projects significantly impact their local area because they are tailored to the region's unique needs. Still, because of the Cquin network, they can also be shared and workshopped to build a national program or adapt to other areas. ICAP and Cquin represent a pinnacle of how to start with differentiated service and attain success within a complex epidemic, but then offer space to scale up the interventions to make the impact global.

Discussion

Schools of public health could form AHSSs to cover all aspects of the discovery-care continuum and close the research-clinical practice gap. However, ensuring translation into practice takes even more coordination. Completing HIV research and supporting humanitarian and policy-driven HIV organizations is not enough. Information must be disseminated among field workers, and collaboration with the service delivery team must be prioritized to identify topics impacting future research. Also, if the service organization supports the specific intervention and geographical region to which the study applies, research findings may still need to be abandoned. The

best research translation occurs when there is mutual faculty across the continuum, such as how Paul Spiegel's research immediately informs the Momentum program in Mali.

In complex epidemics, more than fulfilling the discovery, care is required. With the interplay of other health conditions and social determinants, improving health outcomes takes context-specific research and differentiated service. There are many high-priority research questions, focusing on an area of expertise, and seeing the findings through the continuum is much more productive and a better allotment of resources than maximizing research output. IeDEA's work to leverage observational data is essential to limit the amount of RCT in SSA and, therefore, step away from the neocolonial tendencies of global health. Their work also enables the findings to be shared closer to real-time, which is valuable in a world where EBM is implemented. However, these research projects lack a connection to current implementation programs. Cquin's network of professionals involved in differentiated service delivery offers a better framework for global change. Such frameworks focus on regional successes with options to scale up to global programs instead of trying to apply global research to a particular context.

Questions remain as to why these functional organizations are not more widespread; potentially, external funding structures hinder the cross-cutting of organizations or individual researchers' workload, which impedes their ability to coordinate the entire continuum. What is clear is that integrating the discovery-care continuum through mutual faculty should be considered when planning research, as this enables key stakeholders to communicate throughout the process. Public health schools should make early decisions that allow them to carry context-specific research through the continuum to be potentially scaled. These guidelines allow for successfully translating research to clinical practice in areas that need intervention most.

These projects exemplify the potential of global health to improve their practices. Functional integration of all

pieces of the discovery-care continuum reforms how global health has used LIMC for research but has yet to invest in long-term healthcare solutions. While following the goals of EBM, careful selection of programs to participate in can ensure context-specific research that makes a true impact instead of imposing Western care practices on LIMC. Additionally, partnering with organizations with local leaders and prioritizing capacity building will improve the intervention's success and help global health distance itself from neocolonialism.

Conclusion

From global health's initiation and exacerbated through EBM, the field has been latently tasked with maintaining inequities. Conducting research focused on helping the most afflicted people and ensuring its translation can rectify this legacy. The research-clinical practice gap must be closed, and academic health science systems must offer an avenue to do so. Current public health schools' research and partnerships offer various frameworks for integrating the sectors of the discovery-care continuum. These case studies depict that mutual faculty is a fruitful way to ensure alignment throughout the continuum and that global health projects should focus on granular solutions when the health concern is complex. To work towards equity, global health must be restructured in a way that begins to account for scale at every step of the research process. Research will continue to get funding even if it does not focus on local concerns or ensure implementation. Healthcare is too complex to operate based on statistics that don't translate to worries. EBM needs to be implemented in concert with interventionist approaches to research. Overall, research needs to scale toward fulfilling the discovery-care continuum. Still, each program should somehow work on scaling up, such as building partnerships where ground organizations can let their data be helpful to those setting research priorities. Further research into more guiding principles for and the barriers to implementing these frameworks at public health schools is

needed.

References

1. Wiafe S, Darrach M. The Global Health Community Must Reckon with Realities of Neocolonialism, Racism, and Racial Trauma in Current Practice. Global Health: Annual Review. 2021;1(6). https://journals.mcmaster.ca/ghar/article/view/2656.
2. Biehl J, Petryna A. *When people come first: critical studies in global health.* Princeton University Press; 2013. Accessed June 7, 2023. https://search-ebscohost-com.libproxy.lib.unc.edu/login.aspx?direct=true&db=nlebk&AN =563768&site=ehost-live.
3. Robinson T, Bailey C, Morris H, et al. Bridging the research–practice gap in healthcare: a rapid review of research translation centres in England and Australia. *Health Res Policy Sys.* 2020;18(117). https://doi.org/10.1186/s12961-020-00621-w.
4. Kerridge I. Ethics and EBM: acknowledging bias, accepting difference and embracing politics. *J Eval Clin Pract.* 2010;16: 365-373. https://doi-org.libproxy.lib.unc.edu/10.1111/j.1365-2753.2010.01412.x.
5. Worsham C, Jena AB. The art of evidence-based medicine. *Har Bus Rev*. 2019. Accessed July 21, 2023. https://hbr.org/2019/01/the-art-of-evidence-based-medicine.
6. Tageja N. Bridging the translation gap – new hopes, new challenges. *Fundam Clin Pharmacol*. 2011;25:163-171. https://doi-org.libproxy.lib.unc.edu/10.1111/j.1472-8206.2010.00903.x.
7. Forland F, Rohwer A, Klatser P, et al. Strengthening evidence based healthcare in Africa. *BMJ Evid Based Med*. 2013. https://doi.org/10.1136/eb-2012-101143.
8. Mehta U, Kalk E, Boulle A, et al. Pharmacovigilance: a public health priority for South Africa. *S Afr Health Rev*. 2017;2017:125-133. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5708547/.
9. Edelman A. Can academic health science centers drive translational research to meet health needs in low-income countries? *Global Voices World Bank & IMF Delegation.* 2015. https://static1.squarespace.com/static/56635f09e4b078c56b4d110b/t/5beba2610ebbe88db fe87eba/1542169191172/Final+Research+Paper+-+Alex+Edelman+copy.pdf.
10. Dzau VJ, Ackerly DC, Sutton-Wallace P, et al. The role of academic health science systems in the transformation of medicine. *Lancet.* 2010;375(9718):949-53. https://doi.org/10.1016/S0140-6736(09)61082-5.
11. Park B, Frank B, Likumahuwa-Ackman S, et al. Health Equity and the Tripartite Mission: Moving From Academic Health Centers to Academic-Community Health Systems. *Acad Med*. 2019;94(9):1276-1282. https://doi.org/10.1097/ACM.0000000000002833.
12. Michener L, Cook J, Ahmed SM, et al. Aligning the goals of community-engaged research: Why and how academic health centers can successfully engage with communities to improve health. *Acad Med*. 2012;87(3):285-291. https://doi.org/10.1097/ACM.0b013e3182441680.
13. Sustainable Development Goal 3. United Nations. Accessed August 2, 2023.
14. Rodney W. *How Europe underdeveloped Africa*. Black Classic Press; 2012. https://ebookcentral-proquest-com.libproxy.lib.unc.edu/lib/unc/detail.action?docID=9916 51&pq-origsite=summon. Accessed July 6, 2023.
15. Richardson ET. *Epidemic illusions: On the coloniality of global public health*. MIT Press; 2020. https://ebookcentral-proquest-com.libproxy.lib.unc.edu/lib/unc/detail.action?docID=6416 764. Accessed July 6, 2023.
16. Mlambo A. Western social sciences and Africa: The domination and marginalisation of a continent. *Afr Sociol Rev*. 2006;*10*(1): 161–179. http://www.jstor.org/stable/afrisocirevi.10.1.161.
17. Crane J. Unequal 'partners': AIDS, academia, and the rise of global health. *Behemoth.* 2010;3(3):78-97. https://www.researchgate.net/publication/49616454_Unequal_'Partners'_AIDS_Academi a_and_the_Rise_of_Global_Health.
18. Gautier L, Sieleunou I, Kalolo A. Deconstructing the notion of "global health research partnerships" across Northern and African contexts. *BMC Med Ethics*. 2018;19(1):49. Published 2018 Jun 15. https://doi.org/10.1186/s12910-018-0280-7.
19. Timmermans S, Mauck A. The promises and pitfalls of evidence-based medicine. *Health Aff*. 2005;24(1). https://doi.org/10.1377/hlthaff.24.1.18.
20. Kolodziej MA. Does evidence-based medicine really reduce costs?. *Oncology (Williston Park).* 2011;25(3):214-218.
21. Birbeck GL, Wiysonge CS, Mills EJ, et al. Global health: the importance of evidence-based medicine. *BMC Med.* 2013;11:223. https://doi.org/10.1186/1741-7015-11-223.
22. Galbraith K, Ward A, Heneghan C. A real-world approach to evidence-based medicine in general practice: a competency framework derived from a systematic review and Delphi process. *BMC Med Educ*. 2017;17(1):78. Published 2017 May 3. https://doi.org/10.1186/s12909-017-0916-1.
23. van Baalen S, Boon M. From EBM to epistemological responsibility. *J Eval Clin Pract.* 2015;21: 433-439. https://doi-org.libproxy.lib.unc.edu/10.1111/jep.12282.
24. Kamath S, Guyatt G. Importance of evidence-based medicine on research and practice. *Indian J Anaesth*. 2016;60(9):622-625. https://doi.org/10.4103/0019-5049.190615.
25. Basham C, Billings E, El Rifay A, et al. Designing and validating a One Health Research Translation Framework through literature-based case studies in Egypt. *One Health*. 2022;15. https://doi.org/10.1016/j.onehlt.2022.100454.
26. Crowther H, Lipworth W, Kerridge I. Evidence-based medicine and epistemological imperialism: narrowing the divide between evidence and illness. *J Eval Clin Pract*. 2011;17: 868-872. https://doi-org.libproxy.lib.unc.edu/10.1111/j.1365-2753.2011.01723.x.
27. Representatives of the Global Familial Hypercholesterolemia Community, Wilemon KA, Patel J, et al. Reducing the Clinical and Public Health Burden of Familial Hypercholesterolemia: A Global Call to Action. *JAMA Cardiol*. 2020;5(2):217-229. https://doi.or/\g/10.1001/jamacardio.2019.5173.
28. Burgess RA, Jeske N, Rasool S, et al. Exploring the impact of a complex intervention for women with depression in contexts of adversity: A pilot feasibility study of COURRAGE-plus in South Africa. *Int J Soc Psychiatry*. 2022;68(4):873-880. https://doi.org/10.1177/00207640211010203.
29. Yangui F, Touil A, Antit S, Zakhama L, Charfi MR. COPD prevalence in smokers with stable ischemic heart disease: A cross-sectional

study in Tunisia. *Respir Med*. 2021;179:106335. https://doi.org/10.1016/j.rmed.2021.106335.
30. A and B Recommendations. US Preventive Services Task Forces. Accessed August 17, 2023. https://www.uspreventiveservicestaskforce.org/uspstf/recommendation-topics/uspstf-a-an d-b-recommendations.
31. Subramanian S, Gakunga R, Kibachio J, et al. Cost and affordability of non-communicable disease screening, diagnosis and treatment in Kenya: Patient payments in the private and public sectors. *PLoS One*. 2018;13(1):e0190113. https://doi.org/10.1371/journal.pone.0190113.
32. Carley S, Horner D, Body R, Mackway-Jones K. Evidence-based medicine and COVID-19: what to believe and when to change. *Emerg Med J*. (2020);37(9):572-575. https://doi.org/10.1136/emermed-2020-210098.
33. Misiri HE. Achieving good health and well-being in Africa by 2030 using multi-state models, survival analysis, statistical methods for evidence-based medicine, diagnosis and determination of risk factors. *Stat J IOAS*. 2020;36. https://doi.org/10.3233/SJI-200712.
34. Rogers WA. Evidence based medicine and justice: a framework for looking at the impact of EBM upon vulnerable or disadvantaged groups. *J Med Ethics*. 2004;30(2):141-145. doi:10.1136/jme.2003.007062.
35. Vaidya SS, Guo JJ, Heaton PB, Steinbuch M. Overview and comparison of postmarketing drug safety surveillance in selected developing and well-developed countries. *Drug Inf J*. 2010;44(5):519-533.
36. Lawrence DS, Hirsch LA. Decolonising global health: transnational research partnerships under the spotlight. *Int Health*. 2020;12(6):518–523. https://doi.org/10.1093/inthealth/ihaa073.
37. Pratt B, Loff B. A framework to link international clinical research to the promotion of justice in global health. *Bioethics*. (2012);28: 387-396. https://doi-org.libproxy.lib.unc.edu/10.1111/bioe.12009.
38. Harper A, Pratt B. Combatting neo-colonialism in health research: what can aboriginal health research ethics and global health research ethics teach each other? *J Empir Res Hum Res Ethics*. 2022;17(4):431-454. https://doi.org/10.1177/15562646211058253.
39. Edelman A, Taylor J, Ovseiko PV, Topp SM. The role of academic health centres in improving health equity: a systematic review. *J Health Organ Manag*. 2018;32(2):279-297. https://doi.org/10.1108/JHOM-09-2017-0255.
40. Alving B, Dai K, Chan SHH. *Translational medicine - What, why and how: an international perspective*. S. Karger AG; 2012. https://ebookcentral-proquest-com.libproxy.lib.unc.edu/lib/unc/detail.action?docID=3016 584.
41. Ackerly DC, Udayakumar K, Taber R, Merson MH, Dzau VJ. Perspective: global medicine: opportunities and challenges for academic health science systems. *Acad Med*. 2011;86(9):1093-1099. https://doi.org/10.1097/ACM.0b013e318226b455.
42. Delaney B, Moxham J, Lechler R. Academic health sciences centres: an opportunity to improve services, teaching, and research. *Br J Gen Pract*. 2010;60(579):719-720. https://doi.org/10.3399/bjgp10X532620.
43. van de Wijgert J. Academic health science systems. *Lancet.* 2010;375(9728):1782. https://doi.org/10.1016/S0140-6736(10)60807-0.
44. Wartman SA, Steinberg MJ. The role of academic health centers in addressing social responsibility. *Med Teach*. 2011;33(8):638-642. doi:10.3109/0142159X.2011.590249.
45. World Health Organization. Trends in maternal mortality 2000 to 2020: estimates by WHO, UNICEF, UNFPA, World Bank Group and UNDESA/Population Division. World Health Organization: Sexual and Reproductive Health Research; 2023. Accessed August 7, 2023. https://www.who.int/publications/i/item/9789240068759.
46. Yovo E. Challenges on the road to achieving the SDG 3.2 targets in resource-limited settings. *Lancet*. 2022;10(2):157-158.
47. Newborn mortality. World Health Organization. Updated January 28, 2022. Accessed August 2, 2023. https://www.who.int/news-room/fact-sheets/detail/levels-and-trends-in-child-mortality-re port-2021.
48. Breastfeeding within an hour after birth is critical for saving newborn lives. World Health Organization. Updated July 31, 2018. Accessed August 2, 2023. https://www.who.int/news/item/31-07-2018-3-in-5-babies-not-breastfed-in-the-first-hour of-life.
49. Gillings Humanitarian Health Initiative. Gillings School of Global Public Health. Accessed August 3, 2023. https://sph.unc.edu/global-health/gillings-humanitarian-health-initiative/
50. Breastfeeding advocacy toolkit. The Global Breastfeeding Collective. Updated June 19, 2023. Accessed August 3, 2023. https://www.globalbreastfeedingcollective.org/breastfeeding-advocacy-toolkit?f%5B0%5 D=local_terms_facet_1%3A9b83c1e3-85ba-4a87-b94b-e8126318c509#in-page-search.
51. About Core Group. Core Group. Accessed August 3, 2023. https://coregroup.org/about-core-group/#our-impact.
52. Technical Assistance. Carolina Global Breastfeeding Initiative. Accessed August 3, 2023. https://sph.unc.edu/cgbi/technical-assistance/.
53. Infant and young child feeding counseling: an integrated course. Global Breastfeeding Collective. 2021. Accessed August 7, 2023. https://www.globalbreastfeedingcollective.org/reports/infant-and-young-child-feeding-co unselling.
54. Publications. Carolina Global Breastfeeding Initiative. Accessed August 3, 2023. https://sph.unc.edu/cgbi/publications/.
55. Tomori C, Gribble K, Palmquist AEL, Ververs MT, Gross MS. When separation is not the answer: Breastfeeding mothers and infants affected by COVID-19. *Matern Child Nutr*. 2020;16(4):e13033. https://doi.org/10.1111/mcn.13033.
56. About. PMA. Accessed August 3, 2023. https://www.pmadata.org/about.
57. Selected Projects. Johns Hopkins Bloomberg School of Public Health. Accessed August 3, 2023. https://publichealth.jhu.edu/practice/in-the-field.
58. What We Do. USAID Momentum. Accessed August 3, 2023. https://usaidmomentum.org/what-we-do/.
59. Bell SO, OlaOlorun F, Shankar M, et al. Measurement of abortion safety using community-based surveys: Findings from three countries. *PLoS ONE*. 2019;14(11): e0223146. https://doi.org/10.1371/journal.pone.0223146.
60. Bell SO, Omoluabi E, OlaOlorun F, et al. Inequities in the incidence and safety of abortion in Nigeria. *BMJ Global Health*. 2020;5:e001814.
61. Bell SO, Sheehy G, Hyacinthe AK, Guiella G, Moreau C. Induced abortion incidence and safety in Côte d'Ivoire. *PLoS ONE*. 2020;15(5): e0232364. https://doi.org/10.1371/journal.pone.0232364.
62. Bell SO, Shankar M, Ahmed S, et al. Postabortion care availability, facility readiness and accessibility in Nigeria and Côte d'Ivoire. *Health Policy and Planning*. 2021;36(7):1077–1089. https://doi.org/10.1093/heapol/czab068.
63. McMahon HV, Karp C, Bell SO, et al. Availability of postabortion care services in Ethiopia: Estimates from a 2020 national sample of public facilities. *Contraception X*. 2022;4. https://doi.org/10.1016/j.conx.2022.100087.
64. The Democratic Republic of Congo. USAID Momentum. Accessed August 3, 2023. https://usaidmomentum.org/where-we-work/drc/.
65. Mali. USAID Momentum. Accessed August 3, 2023. https://usaidmomentum.org/resource/west-africa-regional-reference-brief/.
66. Humanitarian-Development Nexus. Johns Hopkins Center for Humanitarian Health. Accessed August 7, 2023.

http://hopkinshumanitarianhealth.org/research/projects/humanitarian-development-nexus momentum-integrated-health-resilience.

67. Blanchet K, Ramesh A, Frison S, et al. Evidence on public health interventions in humanitarian crises. *Lancet.* 2017;390(10109):2287-2296. doi:10.1016/S0140-6736(16)30768-1.
68. Spiegel PB, Burkle FM Jr, Dey CC, Salama P. Developing public health indicators in complex emergency response. *Prehosp Disaster Med*. 2001;16(4):281-285. doi:10.1017/s1049023x00043430.
69. Ataullahjan A, Gaffey MF, Sami S, et al. Investigating the delivery of health and nutrition interventions for women and children in conflict settings: a collection of case studies from the BRANCH Consortium. *Confl Health*. 2020;14:29. Published 2020 May 27. doi:10.1186/s13031-020-00276-y.
70. Nash D, Yotebieng M, Sohn AH. Treating all people living with HIV in Sub-Saharan Africa: a new era calling for new approaches. *J Virus Erad*. 2018;4(Suppl 2):1-4. Published 2018 Nov 15.
71. Hargreaves JR, Pliakas T, Hoddinott G, et al. HIV Stigma and Viral Suppression Among People Living With HIV in the Context of Universal Test and Treat: Analysis of Data From the HPTN 071 (PopART) Trial in Zambia and South Africa. *J Acquir Immune Defic Syndr*. 2020;85(5):561-570. doi:10.1097/QAI.0000000000002504.
72. Jones HS, Floyd S, Stangl A, et al. Association between HIV stigma and antiretroviral therapy adherence among adults living with HIV: baseline findings from the HPTN 071 (PopART) trial in Zambia and South Africa. *Trop Med Int Health*. 2020;25: 1246-1260. https://doi.org/10.1111/tmi.13473.
73. Meka AFZ, Billong SC, Diallo I, Tiemtore OW, Bongwong B, Nguefack-Tsague G. Challenges and barriers to HIV service uptake and delivery along the HIV care cascade in Cameroon. *Pan Afr Med J*. 2020;36:37. Published 2020 May 27. doi:10.11604/pamj.2020.36.37.19046.
74. World Health Organization. Updated recommendations on service delivery for the treatment and care of people living with HIV. World Health Organization; 2021. Accessed August 7, 2023. https://www.differentiatedservicedelivery.org/wp-content/uploads/2022/05/97892400235 81-eng-1-1.pdf.
75. Ford N, Penazzato M, Vitoria M, et al. The contribution of observational studies in supporting the WHO 'treat all' recommendation for HIV/AIDS. *J Virus Erad*. 2018;4(Suppl 2):5-8. Published 2018 Nov 15.
76. Bulstra CA, Hontelez JAC, Otto M, et al. Integrating HIV services and other health services: A systematic review and meta-analysis. *PLoS Med*. 2021;18(11):e1003836. https://doi.org/10.1371/journal.pmed.1003836.
77. International epidemiological Database to Evaluate AIDS Overview. Eunice Kennedy Shriver National Institute of Child Health and Human Development. https://www.nichd.nih.gov/research/supported/iedea#:~:text=IeDEA%20is%20an%20international%20research,globally%20diverse%20HIV%2FAIDS%20data.
78. Tanzania. Jhpiego. Accessed August 7, 2023. https://www.jhpiego.org/countries-we-support/tanzania/.
79. Maruyama H, Franks J, Laki D, et al. Bringing HIV services to key populations and their communities in Tanzania: from pilot to scale. *J Int AIDS Soc*. 2021;24:e25718. https://doi.org/10.1002/jia2.25718.
80. Shelley KD, Frumence G, Mpembeni R, et al. Can volunteer community health workers manage multiple roles? An interrupted time-series analysis of combined HIV and maternal and child health promotion in Iringa, Tanzania. *Health Policy Plan*. 2018;33(10):1096-1106. https://doi.org/10.1093/heapol/czy104.
81. An SJ, George AS, LeFevre A, et al. Program synergies and social relations: implications of integrating HIV testing and counseling into maternal health care on care seeking. *BMC Public Health*. 2015;15:24. https://doi.org/10.1186/s12889-014-1336-3.
82. Shelley KD, Mpembeni R, Frumence G, et al. Integrating Community Health Worker Roles to Improve Facility Delivery Utilization in Tanzania: Evidence from an Interrupted Time Series Analysis. *Matern Child Health J*. 2019;23(10):1327-1338. https://doi.org/10.1007/s10995-019-02783-8.
83. Our Research: Tanzania. Johns Hopkins Bloomberg School of Public Health. Accessed August 7, 2023. https://publichealth.jhu.edu/international-center-for-maternal-and-newborn-health/our-res earch/tanzania/.
84. Wada PY, Kim A, Jayathilake K, et al. Site-Level comprehensiveness of care is associated with individual clinical retention among adults living with HIV in International Epidemiology Databases to Evaluate AIDS, a Global HIV Cohort Collaboration, 2000-2016. *AIDS Patient Care and STDS*. 2022;36(9). https://doi.org/10.1089/apc.2022.0042.
85. Mburu C, Njuguna I, Neary J, et al. Mortality and loss to follow-up among adolescents and young adults attending HIV care programs in Kenya. *AIDS Patient Care and STDS*. 2023;37(7):323-331. https://doi.org/10.1089/apc.2023.0019.
86. Boni SP, Horo A, Didi-Kouko-Coulibaly J, et al. Impact of HIV infection on access to cancer care and survival among women with invasive cervical cancer in Côte d'Ivoire: A prospective cohort study. *Int J Gynecol Obstet*. 2023; 00: 1-10. https://doi.org/10.1002/ijgo.14925.
87. Zürcher K, Cox SR, Ballif M, et al. Integrating services for HIV and multidrug-resistant tuberculosis: A global cross-sectional survey among ART clinics in low- and middle-income countries. *PLOS Global Public Health*. 2022;*2*(3):e0000180. http://doi.org/10.1371/journal.pgph.0000180.
88. Reuter A, Furin J. Celebrating choice in the care of people living with drug-resistant tuberculosis. *Lancet.* 2022;400(10362):1489-1491. https://doi.org/10.1016/S0140-6736(22)01899-2.
89. Romo ML, Brazier E, Mahambou-Nsondé D, et al. Real-world use and outcomes of dolutegravir-containing antiretroviral therapy in HIV and tuberculosis co-infection: a site survey and cohort study in Sub-Saharan Africa. *J Int AIDS Soc*. 2022;*25*(7):e25961. http://doi.org/10.1002/jia2.25961.
90. Brazier E, Maruri F, Wester C, et al. Design and implementation of a global site assessment survey among HIV clinics participating in the International epidemiology Databases to Evaluate AIDS (IeDEA) research consortium. *PloS One*. 2023;18(3):e0268167. http://doi.org/10.1371/journal.pone.0268167.
91. Parcesepe AM, Filiatreau LM, Ebasone PV, et al. Gender, Mental Health, and Entry Into Care with Advanced HIV Among People Living with HIV in Cameroon Under a National "Treat All" Policy. *AIDS Behav*. 2021. http://doi.org/10.1007/s10461-021-03328-3.
92. Parcesepe AM, Filiatreau LM, Ebasone PV, et al. Mental health and initiation of antiretroviral treatment at enrolment into HIV care in Cameroon under a national "treat all" policy: a cross-sectional analysis. *J Int AIDS Soc*. 2021;*24* (11). http://doi.org/10.1002/jia2.25842.
93. Parcesepe AM, Remch M, Dzudie, A, et al. Depressive Symptoms, Gender, Disclosure, and HIV Care Stage Among People Living with HIV in Cameroon. *AIDS Behav*. 2021. http://doi.org/10.1007/s10461-021-03425-3.
94. Yotebieng M, Thirumurthy H, Moracco KE, et al. Conditional cash transfers and uptake of and retention in prevention of mother-to-child HIV transmission care: a randomised controlled trial. *Lancet HIV*. 2016;3(2):e85-e93. http://doi.org/10.1016/S2352-3018(15)00247-7.

95. Preventing the spread of infectious diseases. UNC Gillings School of Global Public Health. 2012. Accessed August 7, 2023. https://sph.unc.edu/cphm/the-power-of-prevention/preventing-the-spread-of-infectious-di seases-3/.
96. How Measure Evaluation Supports Dreams. Measure Evaluation. 2016. Accessed August 3, 2023. https://www.measureevaluation.org/resources/publications/fs-16-183.html.
97. Saving lives through the Dreams Program. Centers for Disease Control and Prevention. Accessed August 7, 2023. https://www.cdc.gov/globalhivtb/who-we-are/success-stories/success-story-pages/Dreams -saving-lives.html#:~:text=With%20support%20from%20the%20U.S.,AGYW)%20partic ularly%20vulnerable%20to%20HIV.
98. Who We Are. ICAP at Columbia. Accessed August 3, 2023. https://icap.columbia.edu/who-we-are/.
99. Accelerating Children's HIV/AIDS Treatment. Elizabeth Glaser Pediatric AIDS Foundation. 2018. Accessed August 8, 2023. https://www.pedaids.org/resource/accelerating-childrens-hiv-aids-treatment-act/#:~:text= XXV this%20project%2C%20titled%20the%20Accelerating,identification%2C%20treatment% 2C%20and%20retention.
100. MTCT-Plus End-of-Project Reports. ICAP at Columbia. Accessed August 8, 2023. https://icap.columbia.edu/tools_resources/mtct-plus-end-of-project-reports/.
101. Creating Hope. ICAP at Columbia. Accessed August 8, 2023. https://icap.columbia.edu/wp-content/uploads/MTCT-Plus_Short-Narrative_FINAL.pdf.
102. Where We Work. ICAP at Columbia. Accessed August 8, 2023. https://icap.columbia.edu/where-we-work/rwanda/.
103. The CQUIN Network: What we do. ICAP at Columbia. Accessed August 8, 2023. https://cquin.icap.columbia.edu/about-cquin/what-we-do/.
104. Differentiated TB/HIV Services. ICAP at Columbia. Accessed August 8, 2023.https://cquin.icap.columbia.edu/network-focus-areas/differentiated-tb-hiv-services/.

Meet the Author:

Mariana Chavez Guerrero attends the University of North Carolina at Chapel Hill with a double double-major in Psychology and Human Development and Family Sciences. She is passionate about promoting educational equity and social justice through the intersection of human development, social psychology, and education. Mariana also serves as a Leadership Fellow for LatinxEd, a nonprofit promoting educational equity for Latinx youth in North Carolina. Through the Moore Undergraduate Research Apprentice Program, under the guidance of Dr. Malissa Alinor, Mariana researched the experiences that influence the development of college students' ethnic and racial identity. In her free time, Mariana enjoys playing the piano, exploring new places, and doing escape rooms. Mariana is currently studying abroad in Spain in order to improve her Spanish and later use it to reach and uplift more members of her community in North Carolina and beyond. She is diving deeper into her research interests through the McNair Scholars Program and later hopes to continue fulfilling her academic endeavors in graduate school.

Space and Self: An Ecological Analysis of Ethnic-Racial Identity Development

Abstract:

As the population of America continues to diversify, it is becoming increasingly important to understand the development of race and ethnicity across the lifespan, especially when these factors influence the health of minoritized individuals. Most research conducted on ethnic-racial identity (ERI) focuses on psychosocial outcomes from the individual's developmental journey, such as mental health, psychosocial adjustment, and academic success. In this paper, I shift the focus from how the individual can be better equipped for healthy adjustment and instead examine the influence that external institutions and communities play in ERI development. I analyzed 20 interviews from 10 Asian Americans and 10 Black Americans between the ages of 18 and 23 years old and focused on questions related to self-identification and moments where they became aware of their role as a member of their racial group. I found three main themes: 1) most participants became aware of their ERI in school settings, 2) the second most common place where participants gained awareness was at home through conversations with family, media, the practice of traditions/customs, or neighborhood interactions, and 3) sports and religious communities were also a space for ERI realization. This examination is essential because it covers a gap in the literature while providing a framework to understand how institutions may affect the wellbeing of diverse youth. Ultimately, this work outlines implications and suggests a collective effort towards promoting healthy ERI development.

Key words: ethnic-racial identity, development, environment, schools, community

Towards Preventative Interventions

As the population of America continues to diversify, it is becoming increasingly important to fully understand the formative roles of race and ethnicity in development across the lifespan. This is especially true when these factors influence the health of minoritized individuals (Umaña-Taylor & Rivas-Drake, 2021). According to the 2020 U.S. Census, the non-white and/or Hispanic population has increased from 36.3% in 2010 to 42.2% in 2020. Moreover, the chance that two individuals picked at random will be from different racial or ethnic groups has gone from 54.9% in 2010 up to 61.1% in 2020 (US Census, 2021), which highlights the increasing diversity of the US over the last decade. This statistic contextualizes the need to better understand the factors that influence identity development - especially when it involves race and ethnicity. Despite this increasing diversity, the U.S. continues to struggle with perpetuating ethnic-racial inequities. Through systemic and institutional means, marginalized communities are affected through unequal access to resources and opportunities, bias, and direct experiences with discrimination. The reality of systemic racism is then showcased through things like poorer mental and physical health, lower academic achievement, and opportunity gaps among youth of color (Umaña-Taylor & Rivas-Drake, 2021). It is important to research ways to promote resilience and more positive developmental experiences in these communities given their impact on health and achievement. In doing this, cycles of inequity can be disrupted by working to reduce ethno-racially based disparities.

According to Social Identity Theory, having a strong positive attachment to one's social group is correlated with having better self-image, mental health, and academic outcomes (Cheon et al., 2020). Within the U.S. context, social groups are often created on the basis of race and ethnicity. Therefore, it's important to investigate how ethnic-racial identity

(ERI) can be most safely developed given its impact on people's health and behavior.

Ethnic-racial identity (ERI) is most commonly defined as a construct that sees individuals' identities as being shaped by both the racialization of their group within their social context and their ethnic features/ancestral heritage which include things like cultural traditions and language (Umaña-Taylor, 2011). Through a metanalysis on the outcomes of ethnic and racial identity development in adolescence, Rivas-Drake (2014) found that most studies show ethnic racial identities (ERIs) are linked to psychosocial functioning - such that positive feelings towards one's ERI was associated with positive psychosocial adjustment. Psychosocial adjustment refers to the ways individuals adapt their needs to changes in their environments, so positive adjustment would indicate a healthy transition into a new environment. The analysis also found that more positive ERIs were positively associated with students' academic success (Rivas-Drake et al., 2014). Moreover, positive ERI was found to be a protective factor against future experiences with racial discrimination, such that ERI resolution served as a buffer against the negative effects of ethno-racially based risk (Umaña-Taylor & Rivas-Drake, 2021).

Most of the research being conducted on ERI focuses on psychosocial outcomes (i.e., "effect-based literature") from the individual's developmental journey. This is an issue because by focusing only on the effects that an environment may have on the individual, these studies fail to consider how the environments themselves could change in order to prevent negative outcomes. By looking at the environmental contexts (institutional atmosphere and social interactions), I build a framework that allows for the construction of preventative strategies to promote healthy ERI development, rather than responsive ones.

In this paper, I shift the focus from how the individual can be better equipped for

healthy adjustment into diverse environments and instead examine the role that contextual mechanisms and social ecologies play in ERI development. By examining this, I cover a gap in the more effect-based literature and find the environmental contexts that require the most attention given their impactful roles in influencing ERI development. Thus, I not only analyze the role of various communities and institutions in youth's development through the analysis of interviews among Black and Asian Americans, but I also outline implications and make suggestions on how working towards healthy ERI development is a large collective effort– not just an individual one. **LITERATURE REVIEW**

In focusing on the way environmental contexts (e.g., communities, institutions) affect identity development, I ground my work on Bronfenbrenner's ecological systems theory (1992). This model shows the individual as the center-most circle in a set of concentric circles. Moving from the inside out, most near to the individual is the microsystem, composed of direct interactions with environmental contexts like home, school, and neighborhoods. Then is the mesosystem, composed of interactions *between* things like home and school or home and parent's work, followed by the exosystem, which is composed of indirect influences to the individual such as mass media and community services. Lastly, in the outermost ring lies the macrosystem, which is composed of larger cultural values, beliefs, and laws (Bronefenbrenner, 1992). This model emphasizes that individuals influence the environments and communities they are placed in, as well as that these environments and communities also influence the individual. This theory is applicable to the overall health of the individual, and thus encapsulates the environmental contexts in which ERI develops.

More specific models have been constructed for the developmental stages and

processes of ERI rather than the more broadly defined wellbeing that Bronfenbrenner's model displays. One such model, Sue & Sue's Racial/Cultural Identity Development Model (2013), outlines the five stages of ERI development: the conformity stage, in which ERI is denied in order to blend in with "American" culture; the dissonance stage, in which individuals become aware of the conflict between "americanness" and being a member of their ethnic/racial group; the resistance and immersion stage, in which exploration of one's own ERI begins through the denial of "American" culture; the introspection stage, in which individuals become aware of their exploration focusing only on their ethnic/racial culture and begin exploring others while learning to trust the dominant society; and finally the integrative awareness stage, which showcases a balanced identity in which there are high levels of membership within one's racial/ethnic group and within the larger "American" society (Sue & Sue, 2013) .

Although the model mimics a lot of the processes of development, I find the linear presentation to be limiting. By drawing on Brofenbrenner's model, I propose that it may be more accurate to consider the "stages"' of the Racial/Cultural Identity Development Model as categories. These processes do not have to be linear, they can occur at the same time, in different orders, under different contexts. The individual's identity is not fixed in a specific place with the same social ecologies for very long periods of time - different environmental contexts, such as home life versus school life produce a fluidity in one's ability to move through these "stages".

My conceptualization of ERI development will follow the idea that Ethnic-Racial identity is multidimensional (Umaña-Taylor et al., 2014). This sees ERI as having both content and process dimensions. Content dimensions are composed of thoughts and feelings related to ERI. These dimensions include affirmation (the individual's own feelings towards their group), public

regard (how the individual thinks others feel about their social group), centrality (how important the individual feels that their ERI is to defining their whole identity), and salience (how important the individual thinks their ERI is in a given situation). Process dimensions include exploration (engaging in activities, conversations, and/or learning more about their ERI) and resolution/commitment (how certain individuals feel about their ERI and what it means for their sense of self). I use this conceptualization to analyze how these dimensions are present in different contexts and how they relate to the emergence of ERI.

The multidimensional approach to ERI was inspired by one of the first models of racial identity, the Multidimensional Model of Black Identity (MMBI), constructed by Dr. Robert M. Sellers (Sellers et al., 1998). This model established the content dimensions previously stated and asserted that they are linked to adjustment and thus will depend on the situational context (Sellers et al., 1998). Although Sellers made this assertion, little research has focused on what the variation between these contexts actually looks like - which is one of the aims of this study. While the MMBI recognizes the impact of societal forces on developing racial identity, it primarily emphasizes the individual's own construction of their racial identity as the most important indicator of racial identity. I don't refute this claim, but rather center the conversation on an ecological analysis of identity since the MMBI focused on an individual's self-construction. Furthermore, I also use these concepts on a sample that includes both Black and Asian Americans.

In accordance with constantly changing environments influencing ERI, one study found that school transitions have an effect on the exploration of and commitment to one's ERI. In a transition from middle school to high school, it was found that adolescents' commitment increased while exploration remained the same (Syed & Azmitia, 2009). This finding is

important because it shows how the environment can make one's ERI fluctuate; however, such transitions may occur through time (going up a grade) as well as space (transitions from school to home). Therefore, places that continuously cater to youth who are undergoing ERI development must be equipped to foster a safe and open space for healthy adjustment.

Previous work on an ecological model of racial identity was focused on biracial siblings. This project revealed four types of experiences that influenced racial identity: hazing, family dysfunction, increased racial integration in society's structure, and the presence of other salient identities (Root. M.P.P., 1998). Although this study has similar aims in understanding experiences that shape racial identity - its sample is only inclusive of biracial siblings, and doesn't focus on a particular age group (ages ranged from 18 to 40 years old). Further, it delves into a traumatogenic impact on one's identity process. My study will look specifically at young adults during an important transition period while also considering race *and* ethnicity. Although experiences with racial discrimination are mentioned, they will not be the focus of the study unless explicitly stated by the participants.

In developing interventions, one of the biggest projects has been the Identity Project (Umaña-Taylor & Douglass, 2017). This project targeted curriculum in order to work *with* youth and their community to increase salience, understanding, and exploration of ethnic heritage of themselves and others while also clarifying any misconceptions on self-expression. This intervention was primarily implemented in schools that had high levels of ethnic diversity. **METHODOLOGY**

The data for this project consists of 34 in-depth interviews of Black and Asian American college students aimed at examining their affective experiences with racial discrimination. These interviews were carried out by trained undergraduate students in exchange for class credit. It's

also important to note that these students - both the interviewer and the interviewee - were located in a predominantly white public university in the southeastern United States. Each undergraduate researcher interviewed two peers with the same racial background in order to minimize social desirability bias – in other words, changing answers in order to appear better to others/feel good about oneself (Larson, 2018). Interviews were also the most ideal way to conduct this research because they exhibit nuances through narrative that otherwise would not be accounted for through quantitative data collection methods. The participant's data was kept confidential through the use of chosen pseudonyms rather than their real names. The interviews lasted from 45 minutes to an hour and were audio-recorded for later transcription by the interviewer. For this research, I looked more closely at 20 of these interviews, composed of 10 Black and 10 Asian American young adults (Table 1) and analyzed commonalities in the communities and places that influenced their ERI development.

Demographics		
	Total	10
	Black	
Male	4	Asian
Female	2 6	
	6	
Average Age	20 years	21 years 10*

Table 1. Demographic data of participants. *Gender data was missing for two Asian American participants.

Coding

I used a inductive approach to read through these interviews and focused on four essential questions:

1. What is your race and/or ethnicity?

2. Why do you identify in that way?

3. Tell me about the time you realized you were a member of your racial group. 4. How did that situation make you feel?

The first question reveals participants' self-identification while the second reveals the more specific influences to their ERI, such as socialization. The third question was important because it allowed me to see *where* these instances of ERI emergence/awareness occurred and the fourth reveals how it impacted their feelings about being a member of their racial group. Since the third interview question was specifically asking about the *first* time participants became aware of their ERI, this means they don't have the protective factor that others with resolved ERIs do. Because of this, these participants were especially vulnerable to experiencing negative feelings as a result of instances where they felt increased salience or were faced with racial discrimination. After categorizing places in which ERI emergence occurred, I then looked at other questions within the interview that helped me form a better understanding of the individual's full developmental context. For example, if the answer to the third question had taken place in school, I would then look at questions such as:

- What was the racial composition of your school like?
- How many teachers did you have that were of a different race than you? ● In what kinds of school-activities did you participate?
- What was the racial composition of your main friend-group?

The sample's age ranged from 18-23 years old, with the average age being 20 years old. This makes the sample very fit for answering questions regarding ERI. This is because during adolescence, there is more of a group consciousness perspective of ERI, meaning that youth primarily make sense of their ERI relative to the people

around them (Quintana, 1999). This sample selection also aligns with Erickson's stages of psychosocial development in which he asserts that it's during adolescence that individuals gain the "cognitive maturity and social exposure" to explore their own beliefs and thus examine their own identity and its place in their larger social sphere (Erickson, 1968). Since the sample trends towards late adolescence and emerging adulthood, these individuals likely have a newly-found capacity to think about ERI in a more nuanced way than members of younger groups. This is important to the research because it allows the participant's answers to be more analytical of the influences that others have on their identity - and therefore provides a better look into their contextual upbringing within their answers.

RESULTS

"I didn't start thinking about the larger institutional and structural forces that shape my identity and personality and politics and the interplay amongst all of those until I went to college… It wasn't until I got here and started studying these issues and talking about these issues with my friends and reading about them in the newspapers and lots of media that I kind of started to extrapolate my experience to the larger Asian American experience and other minority groups and marginalized groups." Bess, a 21 year-old Chinese-American college student

Bess is one of many examples of the bidirectional relationship between an individual's development and their environment. In this case, peers at her predominantly White college helped her better understand her ethnic-racial identity (ERI). This experience differed vastly from her ERI development at home, where during her early adolescence her father told her, "you think you are American, but everyone else sees you as Chinese". Bess explains this was the first time she became fully conscious of her identity and the lack of control she had over it. Whereas at home her identity invoked feelings of helplessness and isolation, making her think things like "I always wonder if I seem perpetually different… forever foreigner to my friends or if they see

me very much as one of them", arriving at college allowed her to explore and find community through shared experience with other Asian Americans and marginalized groups. By analyzing the communities and places that influence ERI development in stories like Bess', I identified three main themes:

1) The most common place in which people came to first realize their ERI happened in schools through relationships with their peers or observation of the dominant culture 2) The second most common place in which people came to first realize their ERI happened in the home- through interactions with family members, the media, neighbors, or the practice of customs and traditions

3) There was a smaller trend pointing towards smaller communities such as religion and sport.

Participant	Ethnic-Racial Identity*	Context where ERI Occurred			Positive Experience
		Negative Experience	Neutral Experience	Emergence	
EJ	African American	Home	X		
Logan	African American/Black	Home	X		
Bess	Chinese-American	Home	X		
Ketki	Indian	Home			X

London Kendra	African Home X American/Black African Home X American/Black	
Total Experiences in Home	6	
Lexie	Sri School X	

	Lankan-American	
Austin	Asian Pacific School Islander, Hmong	X
Kayla	African School X American/Black	
Osmosis	Biracial School X	
Margaret	Vietnamese School X	
Chadsworth	Mixed School X	
AK	Indian School X	
Aliya Adam	Indian School X Bengali School X	
Mikah Denise	Jamaican/Black School X African School X American/Black	

Total Experiences in School Sanjana	11 Indian Religion	X
Ian	Afro-Latino Sport X	
S	Indian Sport	X
Total Experiences in Smaller Communities	3	

Table 2. Distribution of ERI emergence settings, ERI was classified based on participants' own self-identification

These communities, relationships, and environments all shape the way these young adults experienced gaining awareness for their ERI. It's important to look at where positive experiences with ERI occur versus negative experiences. Experiences relating to one's' ERI matter because

most of the time, the negative experiences associated with ERI are a result of ethnic-racial discrimination confusion/anxiety from increased saliency and lack of ERI commitment. Ethnic-racial discrimination has been linked with lower mental health, poor academic adjustment, and higher risky behaviors (Umaña-Taylor & Rivas-Drake, 2021). By coding for negative experiences, we can better identify where the most improvement is needed in order to prevent these negative outcomes and lower achievement gaps. Efforts should be made to not simply prevent negative experiences, but to also promote positive ones. For many of these participants, positive experiences were also moments of healthy ERI exploration and/or affirmation. These moments are important because they help guide individuals

towards ERI resolution/commitment, which then has protective effects on future experiences and are associated with higher self esteem, better mental health, and more positive academic outcomes (Cheon et al., 2020).

The Home

From the sample (n=20), 30% of people's first realization of being a member of their racial group occurred in the home (three Asian Americans and four Black Americans). These answers ranged from stories like Bess' to comparisons made from the home to other places, such as in Kendra's case. As a Black American, she reflects: "My neighbors always liked to play in their sprinklers, and I always wanted to play with them, but my mom was like we don't do that. We don't play with water like that, so I feel like that was the first time I kinda like realized like wow, I'm different." Kendra's experience was coded as negative, given that she remembers: "I was probably really upset, you know like oh... this is something different, like this is not... like what do you mean we're different?" On a similar note, London also made note of this difference but rather than comparing her home life to her neighbors', she compared it to the representation

shown on TV. In reference to knowing she was Black, London said she "kind of always knew when [she] would look at the TV and realized that there were people who didn't look like [her]." London described this experience as being "neutral". Although Kendra and London experienced different feelings, their experiences were rooted in increased salience. Their ERI became a more important and apparent part of their identity when playing in the neighborhood or consuming media.

When ERI was first realized in the home, there was a more even distribution of positive and negative experiences. Ketki, a 21 year-old Indian woman who grew up "surrounded by these cultural influences and participating in activities with [her] family" said she felt "nothing but pride for being Indian. I think a lot of my views are influenced by my parents.

They give me a positive outlook on my race". In this case, Ketki is showcasing how the process of ERI exploration and resolution led her to a positive affirmation on her ERI. Two people showcased resilience: EJ, a 19 year-old Black female and Logan, a 23 year-old Black male. For EJ, despite initial negative feelings she says "my mom was strong enough to teach us that stuff is going to make you mad or sad but it's not going to kill you, at least not the verbal stuff, but you just gotta keep moving forward just got to walk away". For Logan, he explained that his mom had always told him "since you're Black, you'll have to work twice as hard", but he only chooses to think about this when he's "busy thinking about how someone else feels about me… I find it's better to like just not do that all… I try not to worry about other people's perception." Logan and EJ's experiences were coded as negative since that aligns more with their initial feelings on the situation. However, I highlight resilience because it shows how these individuals were able to combat the negative effects of racial discrimination to their ERI affirmation and public regard by

either speaking to someone of the same race about their experiences (Ketki) or by choosing to prioritize affirmation over public regard (EJ).

Smaller Communities

Sanjana, an Indian student identified her Hindu religion as the source for the awareness of her ERI, sharing that "growing up my parents always taught me the different parts of my culture by attending prayer sessions…". Two participants (a Black American and an Asian American) identified sports as being key roots for their awareness. For Ian, an Afro-Latino male, playing a predominantly White sport like lacrosse made him aware of the stereotype that "Black kids don't play lacrosse". Conversely S., an Indian male was heavily influenced by joining an Indian dance community in America, which made him feel pride in "being more Indian".

The single experience where ERI was

first realized through a religious context (prayer sessions and traditional dance) was associated with positive feelings. Sanjana said "It was really eye-opening… I find myself reading more books about my culture and asking more questions and researching more. It makes me more excited to learn more things about my culture and I'm proud because I like talking about my culture." In this sentence, Sanjana perfectly embodies how exploration can have positive effects on ERI content dimensions. This is similar to S., who described being involved with dance (exploration) as a positive experience.

When looking at experiences where ERI was first realized through sports, Ian showed resilience by looking at his situation through an asset-based perspective: "they actually kind of make me feel better about myself. More so as like 'wow you're doing something that not everybody does', like you're kind of breaking barriers. So I'd say pretty good". In doing this, he makes his ERI more central to his identity, and reframes public regard.

School

More than half of the sample (55%) indicated that they first became aware of their ERI at school. This group was composed of six Asian Americans and five Black Americans (Table 2.). The school grades of realization ranged from daycare to college. This realization was mainly led by interactions with school peers and an emerging awareness of the difference between the self and the environment. This increased saliency is showcased by Lexie, a Sri Lankan American woman who said, "I ate with my hands at home and stuff, that's not something I would do at school… that was specifically Asian about me. Where like the food we eat at home we don't eat at school, it's just distinct". For Lexie, this realization led her to doing things like pretending her mother's packed lunch was a peanut butter and jelly sandwich rather than a marmite and butter sandwich. This behavior showcases negative affirmation and lower self-confidence.

Four students explained the realization came from a specific interaction with their peers.

For example, Kayla, a Black female shared that in kindergarten a girl said to her “Oh, I don’t know if I could be your friend because I’ve never had a brown friend before”. She said this made her feel weird because of how young she was- “I was confused on why she would think that she couldn’t be my friend because I was ‘brown’”. However, although still in a school environment, this experience can look very different in a college setting. Mikah, a Black Jamaican female shared that growing up in England and Massachusetts, she was only ever surrounded by White people. Thus, upon arriving to Georgia, she says “I was around black people, it was kind of like… I didn’t really act Black and people would tell me that. Like the way I talk, the way I acted, things like that”. Because these comments were coming from in-group members, they pose a threat to Mikah’s self-concept. Unlike public regard, where the individual is concerned with how out-group members perceive their ethnic-racial identity, Mikah’s case is more linked with affirmation. Private-regard is more influential on the individual’s sense of belonging, since it is concerned with how people of their same identity/group are accepting of them. 55% of the total participants recalled experiencing negative feelings such as shame, insecurity, and distress upon realizing they were a member of their racial group. This was showcased in attitudes and behaviors like: “I tried to hide that I was different”, “how would this experience, although it’s been amazing, would have been different if I was white”, and “I was probably really upset… like what do you mean we’re different?”. However, when looking at the distribution of these negative experiences, 63.6% of them occurred in school settings. Furthermore, only one out of all the students who first realized their ERI in a school setting described a positive experience. Austin, a 20 year-old Hmong male realized he was a member of his racial group when he had to take English Speakers of Other Languages(ESOL) classes in elementary

school. When asked to describe his feelings about this experience he said “well I was pretty young so I didn’t really understand anything then. But thinking back to it now, yeah it was good and very enlightening I guess you can say”.

Other participants related to this feeling of confusion, seeing as the school years at the time of realization were mostly during elementary and middle school. However, these feelings were not remembered as being “good and very enlightening”. Rather, they were described in a more distressing way. For example in kindergarten, Kayla, a 21 year old Black American first realized she was a member of her racial group when a girl said to her: “oh, I don’t know if I could be your friend because I’ve never had a brown friend before”. She reflected on her feelings during this moment by saying “It made me feel weird because I didn’t really know what she was talking about… I was confused on why she would think that she couldn’t be my friend because I was ‘brown’”. In this case, Kayla didn’t have a resolved ERI to use a protective factor when encountering this experience with racial discrimination. Similarly Deinse, a 22 year-old Black American, felt confusion upon moving from a majority Black school to a predominantly White one. On the first day of school, she heard her parents being called a racial slur- “At the time I was mostly just confused. I didn’t understand why this was all happening. I was more affected in middle school because by then I had a better understanding of race and racism.” Although most of the literature explains that it is during adolescence that ERI plays a more important role relative to other periods of life (Umaña-Taylor & Rivas-Drake, 2021), my findings suggested that these meaningful experiences occurred prior to adolescence. This is most clearly seen when ERI awareness happened in school because most participants are able to recall the grade they were in. It’s then revealed that ERI awareness typically emerges prior to adolescence, primarily during the elementary and early middle school years.

Two of the students had more neutral

reactions to a realization in school, saying things like “I was probably like okay cool and moved on” and “I feel like it was pretty neutral because I never had a bad experience with someone calling me out for being Asian or anything”. Lastly, two students provided responses that showed resilience. Chadsworth, a 21 year-old biracial male, became aware of his ERI after transferring from a predominantly Black charter school to a predominantly White school for third grade. During this process he remembers “looking around… there were no Black kids in my class who looked like me…” but after finding another biracial friend he said they “really clung together because we were both like mixed. We weren’t just Black; we were mixed. We were like ‘oh cool, you’re just like me’.” 21 year-old Mikah, the Black Jamaican woman previously mentioned, also showcased resilience through embracing her own identity. Although in the beginning she attempted to cope with negative in-group comments by seeking acceptance from the out-group: “when I got to college I tried to fit in with White people”, she overcame this by sharing that “when I met people of my race, I was like ‘okay there’s really no reason for me to be like I got to act this way’ like I am Black.” Both Mikah and Chadsworth overcame the adversity of identity confusion and unstable saliency by finding community with others that shared the same racial identity and provided positive affirmations.

DISCUSSION

My findings demonstrated that ERI development is influenced by communities and institutions. More specifically, by coding the interviewees' response to when they first became aware of their ERI based on their environment and social ecologies, I found that school was the most common place for the emergence of ERI development. Despite school being the place where over half of the participants first realized their ERI, these experiences were overwhelmingly negative. These negative experiences were usually a result of unstable saliency (experiencing significantly different levels of saliency from one environmental

context to the next due to the reliance on the environment and social cues to dictate one's identity), an indicator of lack of commitment, or because of direct experiences with racial discrimination and microaggressions. Thus, to promote more positive experiences with ERI development, preventative efforts should be implemented through communities and curriculum.

Through the interviews conducted, it is clear that positively discussing students' ERI in schools can have positive impacts. The participants were asked if there was a teacher who had a meaningful impact on their life and to describe why. Aliya, a 22-year-old Indian woman who first realized her ERI in school by being teased for the color of her skin and hair, said her White AP Statistics teacher left a meaningful impact on her. She says that "he was always really open. [He] tried to talk to me about my background, my life, my history; so I think that teachers that express interest in you beyond school-related things have always stuck out to me and he's been one of those people." Denise, the 22-year-old Black woman who also first realized her identity in school by transferring to a PWI, recalled that a meaningful teacher to her was her White senior year language arts teacher. She notes: "I really respected her because she was the first White teacher that really acknowledged my race but not in like a 'I'm woke' type of feel but genuinely acknowledged that race used privilege and how that works." Both of these students show the powerful impact that talking about ERI in schools can have. Despite both teachers being of a different race than them (White), they were still able to engage in ERI exploration through conversations surrounding their identity. Positive experiences with ERI can promote healthier and more positive self-concepts.

Discourse on the role of race and ethnicity in institutions is very relevant to today's socio political context- especially when referring to schools. In June of 2023, The Supreme Court struck down the use of affirmative action, a doctrine established in

1965 allowing race to be considered as part of a student's application in college admissions. Throughout the past years, there have also been substantial attacks on the use of "critical race theory" in K-12 public schools, leading to the ban of race education in many districts. These legislative efforts aim to create colorblind school environments by diminishing the salience of ERI, which is a key predecessor to ERI developmental processes (Douglass et al., 2016). Policies banning racial education also make it difficult for teachers to engage in conversations like the ones had with Denise and Aliya. Thus, it is important to look at ways to mitigate the effects of this legislation in order to promote exploration, confidence, and commitment through youth's ERI.

A key pattern from the interviews was that many students became aware of their ERI when it became a more salient identifier. This is consistent with the Multidimensional Model of Black Identity because it supported the claim that salience is a function of the interaction between an individual and their immediate context, so it must change from moment to moment for the same person (Sellers et al. 1998). In this way, I found that ERI development can occur through both time and space, rather than just space as supposed by Syed & Azmita (2009). For example, when moving from a space where the majority of people in ones' environment shared the same ERI to a space where the individual became minoritized, awareness of ERI emerged due to its increased saliency. For a lot of the participants, this also meant that they experienced unstable ERI salience due to the everyday variations of saliency between their home and school life. I propose that instead of expecting children to create resiliency on their own, we can combat unstable saliency by encouraging healthy exploration, commitment, and resolution that can begin in the home.

Unstable salience experiences likely arise from the lack of a strong sense of commitment (Douglass et al., 2016). Once commitment(internal self-concept) to one's

ERI is formed, one can draw on this to inform their salience as opposed to relying on contextual cues. This means confidence can be carried *within* the individual regardless of changes in the environment. Therefore, it is important to build a strong foundation of commitment to encourage experiencing salience in a more consistent way that prevents feelings of anxiety and confusion.

To work around the legislative attacks on ERI curriculum in schools, encouraging this commitment could begin in the home or other smaller communities. As seen in instances like Ketki's, she noted that her parents instilled the confidence in her to be proud of her own Indian culture. When asked how he felt discussing racial topics with his parents, Chadsworth said: "I have always felt more empowered… When I talk about race, I feel like I was using like a close-minded, very emotional state, and then, talking to them, they would bring me back onto a more logical track on how to look at it and perceive the situations. And it always fell back onto you know don't change and always love where you came from and always love who you are." Ketki and Chadworth is an example of how home life can help develop a strong sense of commitment to one's ERI, that can then serve as a protective factor in any possible future encounters with racial discrimination. Friends can have a similar effect. For Margaret, a 21 year-old Vietnamese woman, she noted that having a diverse friend group on social media created "an uplifting type situation where [they] always like encourage each other to post and be proud of who [they] are." Community and familial efforts can encourage ERI exploration and cultural/racial pride to build commitment. This can then serve as a protective factor when placed in school settings where conversations like the ones in the home or with friends would not be possible.

Although community organizing can serve as a strong ecological tool for promoting positive ERI development, this does not mean efforts to make school a

similarly safe space should be stopped. As seen with the Identity Project, curricular interventions can make a positive impact and should be something any school should strive for when considering students' wellbeing and experiences within its walls. When institutional school efforts are paired with community efforts, it opens the potential for everyone to experience ERI development in the most positive and uplifting way.

LIMITATIONS AND FUTURE DIRECTIONS

A key limitation for this study was that the findings are not generalizable due to the small sample size. Thus, a starting point for future studies on ecological perspectives of ERI development can be gathering a larger sample size. After finding that school settings were particularly impactful in dictating when minoritized youth first realized their ERI, future research should look at what specifically about schools are perpetuating these issues. This could be done by looking at things like school climate and curriculum as influential institutional forces. This study also grouped together Asian and Black Americans but did not look at between or within group differences. This is another important avenue for future research to look at because different minoritized identities experience oppression differently. So, although shared experiences were found between these groups it should be noted that there could be differences both between and within each racial group that this research did not take into consideration. Moreover, since this study was not longitudinal, there is no way to actually measure how positive experiences shaped ERI commitment and were consequently used as protective factors. Despite this, it is an assumption that is based on the rigorous literature that has previously established this link.

Lastly, ecological factors can affect any form of identity. It would be interesting to look at things like gender identity through a similar lens of the study and see how the findings may vary when thinking about initial awareness.

CONCLUSION

This study provides insight into *where*

experiences of emerging ERI awareness and development occur, *how* (through which mechanisms) these places are triggering the onset of ERI development, and which types of experiences (i.e. positive or negative) are happening in corresponding spaces. ERI has been studied under many different lenses. I assert that it is heavily influenced by the communities and institutions that the individual takes place in through sharing the stories of Black and Asian American young adults in the Southeastern United States. In showcasing their stories, I reveal the responsibility these places have in ERI development and construct collective- and preventative- intervention efforts that may be implemented to promote wellbeing and healthy development. Similar to the Identity Project, this could be accomplished through curricular changes or interventions in schools, but could also be accomplished through community education and organizing efforts.

References

Bronfenbrenner, U. (1992). Ecological systems theory. In R. Vasta (Ed.), *Six theories of child development: Revised formulations and current issues* (pp. 187–249). Jessica Kingsley Publishers.

Cheon, Y.M., Ip, PS., Haskin, M., Yip, T. (2020). Profiles of adolescent identity at the intersection of ethnic/racial identity, american identity, and subjective social status. *Frontiers in Psychology*, 11:959, 1-13.

Douglass, S., Wang, Y., Yip, T. (2016). The everyday implications of ethnic-racial identity processes: exploring variability in ethnic-racial identity salience across situations. *Journal of Youth and Adolescence*, 45, 1396-1411.

Erikson, E. H. (1968). Identity: Youth and crisis. New York: Norton.

Larson, R.B. (2018). Controlling social desirability bias. *International Journal of Market Research*, 61(5), 534-547.

Quintana SM (1994). A model of ethnic perspective-taking ability applied to Mexican-American children and youth. International Journal of Intercultural Relations, 18, 419-448 Rivas-Drake, D., Markstrom, C., Syed, M., Lee, R.M., Umana-Taylor, A.J., Yip, T., Seaton, E.K., Quintana, S., Schwartz, S.J., French, S. (2014) Ethnic and racial identity in adolescence: implications for psychosocial, academic, and health outcomes. *Child Dev*. 85(1), 40-57.

Sue, D. W. & Sue, D. (2013). *Counseling the Culturally Diverse: Theory and Practice*, 6th Edn. New York, NY: John Wiley and Sons, 297.

Syed, M. & Azmitia, M. (2009). Longitudinal trajectories of ethnic identity during the college years. *Journal of Research on Adolescence*, 19(4), 601-624.

Sellers, R.M., Smith, M.A., Shelton, J.N., Rowley, S.A.J., Chavous, T.M. (1998). Multidimensional model of racial identity: a reconceptualization of African American racial identity. *Personality and Social Psychology Review*. 2(1), 18-39.

Umaña-Taylor, A. J. (2011). Ethnic Identity. In S. J. Schwartz, K. Luyckx, & V. L. Vignoles (Eds.), Handbook of identity theory and research (pp. 791– 809). New York: Springer. Umaña-Taylor, A.J. & Douglass, S. (2017) Developing an ethnic-racial identity intervention from a developmental perspective: process, content, and implementation of the identity project. (2017). *Handbook on Positive Development of Minority Children and Youth*. 437-453.

Umaña-Taylor & Rivas-Drake. (2021). Ethic-racial identity and adolescents' positive development in the context of ethnic-racial marginalization: unpacking risk and resilience. *Human Development*, 65(5-6), 293-310.

U.S. Census Bureau. (2021). *2020 U.S. Population More Racially and Ethnically Diverse Than Measured in 2010*. U.S. Department of Commerce. Retrieved June 26, 2023, from data.census.gov.

Meet the Author:

My name is Amie Boakye, and I am originally from Louisville, KY and I am a Sophomore at the University of North Carolina Chapel Hill. I will be graduating a year early with the Class of 2025 and aspire to be a Fulbright and/or Rhodes Scholar. I am 19 years old and I am majoring in psychology, B.S. and political science B.A. with a minor in Spanish and will completely all the requirements for a minor in neuroscience as well as being heavily active in the UNC Chapter of March for Our Lives against gun violence. The MFOL team has had multiple meetings with legislators to try and enact common sense gun policy to protect our communities. I am an active team member of the UNC All-Girl Club Cheer competition team. This semester, I am also conducting a research project with Dr. Isaac Unah on the patterns of prison violence in the United States in the past two decades. Additionally, I advocate heavily for the mental health of students and athletes as someone who has had severe struggles with this area and wellness. Throughout my academic journey, I have written numerous papers and historical analyses on this topic, delving deep into the policies and practices that have shaped our current state of affairs, although this is my first attempt in publication. My connection to this subject extends beyond the confines of academia. As a Black woman from Louisville, Kentucky, my experiences are deeply intertwined with the realities of systemic racism and police violence. The murder of Breonna Taylor, a tragic victim of a "failed drug bust" that was purportedly part of gentrification efforts for the downtown area, shook my community to its core. As a 15-year-old going into my junior year, I had to witness the aftermath firsthand – curfews, the presence of the National Guard, and protests where tear gas was deployed, and rubber bullets were fired. These events heightened my awareness of being a Black individual in America, where simply existing can be fraught with danger and suspicion. These firsthand experiences have fueled my passion for seeking justice and reform within our criminal justice system. My academic pursuits are not merely intellectual exercises; they are driven by a desire to effect real change. I am committed to using my education and knowledge to advocate for a shift away from punitive approaches towards a more rehabilitative model, akin to that of countries like Sweden. I believe that by focusing on rehabilitation and addressing the root causes of crime, we can create a system that promotes healing, redemption, and genuine societal transformation.

Federalism in the Drug Crisis in the 1980s

To what extent did the Reagan administrations war on drugs negatively impact low socioeconomic communities?

The Kernell and Smith textbook points out the question "how can the federal, state, and local governments provide a coordinated effective response?" (e.g. Kernell and Smith 2019, p. 47).[1] In the case of the crisis against drugs and public hysteria against drugs the government utilizes its abilities to cooperate against this issue. However, in cooperating to fix the issue, the federal government was impulsive and unwilling to listen to the opinions of others advising against a drug war. In addition, Kernell and Smith state that "federalism means war", and in the case of this crisis it means a war on drugs (e.g. Kernell and Smith 2019, p.49).[2] Likewise, the Kollman readings make the point that federalism is a response to collective dilemmas and in this case it happened to be somewhat successful in the wrong ways (e.g. Kollman 2019, p. 75-76).[3]

On January 20, 1981, the country brought Ronald Reagan into office on the epitome of hope. During Reagan's presidency, the administration drastically cut state budgets, raised taxes for the lower class, and heightened the gap between the wealthy and poor. On October 14, 1982, Ronald Reagan declared war on drugs. This declaration was based off the public outcry for legislation against the illusive drug issue. At the federal level, Reagan started by expanding many of Nixon's drug war policies that Jimmy Carter, the preceding president, had not touched during his term (e.g. Landmark 2021).[4] The Reagan administration initiated a media campaign aimed at altering public perceptions of substance use and the risk posed by illicit drugs (e.g. Landmark 2021).[5] His harsh stance on drugs led to many pieces of legislation that unintentionally targeted impoverished communities. Still, the Reagan

administration's war on drugs began to negatively impact people with low socioeconomic status at both the federal and state level creating lasting effects on communities with a higher percentage of people of color.

In the 1980s, the public was beginning to view drugs as a threat to society leading them to emulate Reagan's motions to halt the drug issue. In turn, legislative and executive branches moved quickly to pass and fund drug enforcement tactics. Between 2% and 6% of Americans thought substance addiction was the country's "number one epidemic" in 1985, yet, in 1989, the figure had risen to 64% (e.g. Drug Policy Alliance).[6] To pursue the motion to halt drugs, the Reagan administration sought to expand FBI anti-drug spending from $8 million to $95 million. In addition, the administration significantly increased anti-drug funding at the Department of Defense from $33 million in 1981 to $200 million in 1988 (e.g. Trine 2015); (e.g. Thevenot 1997).[7];[8] Through these instances, the media campaign led to a series of anti-drug funding initiatives and garnered support for the anti-drug policies the administration was beginning to put in place. During Reagan's term, large cash grants were provided to law enforcement organizations that were inclined to make drug arrests and offenses top priority. The policy of providing financial incentives to law enforcement organizations to prioritize drug arrests amplified the intensity of the drug war. In addition to giving substantial funds to law enforcement officials as an incentive to enter the drug war, the Pentagon issued more than 1.2 million parts of military weapons to police departments (e.g. Trine 2015).[9] The consequence of this militarization was the erosion of constitutional rights, as law enforcement agencies exploited qualified immunity to carry out aggressive tactics with impunity. Additionally, paramilitary units (SWAT teams) were developed in nearly every major city to fight the drug war (e.g. Trine 2015). SWAT teams served forced, unannounced entry into homes (e.g. Trine 2015).[10] Subsequently, the aggressive and violent nature of these raids not only failed to effectively address drug-related issues but

also contributed to a climate of fear and distrust between law enforcement agencies and the communities they serve; the indiscriminate use of force by SWAT teams, coupled with the legal protections afforded by qualified immunity, created a dangerous environment where innocent lives were needlessly endangered and constitutional rights were routinely violated.

To endorse the War on Drugs, state police forces' resources and funding were increased (e.g. Cooper 2016).[11] Additional, state, and local funding for police departments led to a significant increase in the number of officers on the streets (e.g. Cooper 2016).[12] During the War on Drugs, stop and frisks became increasingly common in the United States (e.g. Cooper 2016).[13] The court often backed the stop and frisks allowing officers to conduct searches of the passenger compartments of cars to search for weapons even if there was no reasonable suspicion such as in the case of *Michigan v. Long* (e.g. Fradella, pg. 58).[14] Further, the court retreated from the idea that suspicion needed to be based on more than just hunches. For example, in *Michigan*

Department of State Police v. Stiz and *Alabama v. White,* the court held the decision that vehicles can be stopped without any evidence or suspicion of impaired driving (e.g. Fradella, pg. 58).[15] The Supreme Court, in addition, ruled in *Whren v. the United States* that an officer's arbitrary reasons for a stop are insignificant to Fourth Amendment evaluation and that the legality of the stop could be decided exclusively through an objective assessment of the entirety of the conditions (e.g. Fradella, pg. 59).[16] Under *Whren* it does not matter if a police officer's intentions were biased if they can provide an "objective" justification for detainment or arrest. This perspective takes "an apartheid approach" to the Fourth Amendment and actively endorses racial profiling of neighborhoods with a high percentage of people of color (e.g. Fradella, pg. 59).

The drug war yielded serious repercussions for African Americans. Due to over-policing, bigger funding for the DOD, anti-drug acts,

military-style operations, and court preceding's the United States began a practice known as mass incarceration. Although mass incarceration was accentuated by Bush and Clinton in the 90s, Reagan's policies enacted the practice making it widespread. Prison numbers have more than tripled since the War on Drugs was declared in 1982 (e.g. Nunn 2002).[17] From 1979 to 1989, the number of African Americans convicted of drug crimes doubled from 22% of all arrests to 42% (e.g. Nunn 2002).[18] During this same period, the overall number of African American arrests for substance abuse offenses increased by over 300 percent, from 112,748 to 452,574 (e.g. Nunn 2002).[19] Perhaps most troubling is the statistic that around 80% of those imprisoned in federal prisons and 60% in state prisons for drug-related convictions are Black or Latino/Hispanic.[20] This is in stark contrast to the demographic composition of the United States, where Black persons make up 13.6% and Latino/Hispanic persons make up 19.1%.[21] This surge in incarceration is particularly alarming when examining the racial disparities within the criminal justice system. This statistic appears even worse when it becomes known that Black and Native Americans are more likely than any other race or nationality to be killed by police forces.[22] This connects the issue of mass incarceration to broader problems within law enforcement and the criminal justice system, suggesting a systemic bias and discriminatory practices.

The increased focus on drug arrests and the deployment of militarized tactics disproportionately affected economically disadvantaged neighborhoods. Poverty, in this context, played a pivotal role in shaping the impact of aggressive law enforcement policies. The over policing of these areas not only intensified the criminalization of poverty but also contributed to a cycle of arrests and incarceration that further marginalized already vulnerable communities. The lack of resources and economic opportunities in impoverished neighborhoods made residents more susceptible to being caught in the crossfire of

aggressive law enforcement tactics. Put simply, poverty limits access to health services, raises social isolation, and has a damaging effect on mental health, enabling addiction to thrive. In this context, addressing the roots of drug related issues requires a comprehensive approach that considers the socio-economic factors contributing to addiction. Simply relying on militarized law enforcement tactics not only fails to address the underlying issues but also perpetuates a cycle of poverty, criminalization, and mistrust between communities and those entrusted with their safety. To build healthier and safer

communities, policies should focus on addressing the social determinants of drug-related issues, including poverty and its associated challenges.[23]

Some historians would argue that Reagan addressed one of the countries' biggest issues. During his time as president, Reagan signed four major crime bills that reduced the high rate of crime being: the Comprehensive Crime Control Act of 1984, Sentencing Reform Act of 1984, Executive Order 12564 of 1986, and Anti-Drug Abuse Act of 1988.[24] The Comprehensive Crime Control Act of 1984 repealed the Bail Reform Act of 1966 and authorized courts to consider dangerousness when setting bail conditions, and establish pretrial detention if necessary (e.g. Congress 1983-84).[25] While this provision aimed to enhance public safety by preventing potentially dangerous individuals from being released on bail, critics argue that it contributed to the problem of pretrial detention and infringed upon the rights of individuals who had not been convicted of a crime. The Sentencing Reform Act of 1984 “reformed the federal sentencing system by dropping rehabilitation as one of the goals of punishment; creating the U.S. Sentencing Commission and charging it with establishing sentencing guidelines; making all federal sentences determinate; and authorizing appellate review of sentences”(e.g. CRS 2009).[26] While this act aimed to promote consistency and fairness in sentencing, critics argue that it may have led to a punitive rather than rehabilitative

approach to criminal justice and contributed to over-incarceration, particularly of nonviolent offenders. The Executive Order 12564 of 1986 "established the goal of a Drug-Free Federal Workplace and made it a condition of employment for all Federal employees to refrain from using illegal drugs on or off-duty"(e.g. SAMHSA 1986).[27] While this order aimed to promote workplace safety and productivity, critics argue that it may have infringed upon employees' privacy rights and disproportionately affected individuals with substance abuse disorders who may require treatment rather than punitive measures. This Federal Anti-Drug Abuse Act of 1988 "intends to prevent the manufacturing, distribution, and use of illegal drugs" (e.g. DOJ 1988).[28] While this act sought to deter drug-related activities and promote public health and safety, critics argue that it exacerbated racial disparities in the criminal justice system, particularly through mandatory minimum sentences for crack cocaine offenses, which disproportionately affected African American communities. Reagan's Administration and congress argued that these bills were for the benefit of the country as they aimed to get drugs off the streets and make neighborhoods safer. Their statements followed the idea that these laws potentially could help Black & African American families facing major drug problems.

Nevertheless, these five acts created immense turmoil for impoverished families. They eliminated parole for most federal prisoners, allowed assets to be confiscated before charges had been filed, increased mandatory minimum sentences, and introduced the death penalty to drug kingpins. The worst part of it all though is the fact that extensive research has shown that incarcerating people for drug-related offenses does not even affect substance abuse rates. Instead, incarceration has been related to an increased risk of a drug overdose circulating a cycle of drug use, arrest, and poverty. Therefore, through these acts, communities of color suffered severe amounts from policies and bills put in place by Reagan. The impact of incarceration on communities with a high percentage of people of color from the war on drugs proved

to be negligible.[29]

The federal system addressed the crisis similar to how a unitary or confederation would handle it. Seeing as the issue arose from public hysteria about the drug issue, state and federal governments felt pressured to create legislation immediately to fix the issue that was not necessarily prevalent. Under pressure, any government has the ability to act irrationally, which is essentially what occurred in the case of the war on drugs. Even with this, there was no such coordination problem that occurred during this crisis and partisan cooperation played a huge role in the drug crisis. There was little division and polarization when writing and passing bills on the topic of this crisis. By working together efficiently, state governments agreed to have a no tolerance policy on drugs. On the judicial side, "judges have a critical role to play in federal drug law reform in light of Congress' long-standing failures to meaningfully change the laws" (e.g. Zunkel and Siegler 2020, p. 1).[30] Although there wasn't action taken by federal judges at the time, in modern days they have the capability to review drug sentencing guidelines and use the Supreme Court's *Kimbrough v. United States* to reevaluate the criminal justice system. Furthermore, while the judicial branch cannot completely fix the pieces of this crisis that Congress has ignored, it "would send a clear message from one co-equal branch of government to another that substantive reform is urgently needed" (e.g. Zunkel and Siegler 2020, p. 1).[31]

In short, the Reagan administration may have caused a small positive change in the crime rate in the United States but overall, their laws and initiatives negatively impacted thousands of impoverished neighborhoods in the United States through both federal and state legislation. As shown above, many of our neighborhoods have been tremendously impacted and are still facing the repercussions of the war on drugs in modern days.

References

"A History of the Drug War." Drug Policy Alliance. Accessed April 22, 2021. https://drugpolicy.org/issues/brief-history-drug-war.

"Anti-Drug Abuse Act of 1988: Public Law 100-690, 100th Congress - Title I: Coordination of National Drug Policy." Anti-Drug Abuse Act of 1988 Office of Justice Programs. Accessed February 10, 2023. https://www.ojp.gov/ncjrs/virtual-library/abstracts/anti-drug abuse-act-1988-public-law-100-690-100th-congress-title-i.

Alexander, Michelle. "The War on Drugs and the New Jim Crow." *Race, Poverty & the Environment* 17, no. 1 (2010): 75-77. Accessed May 17, 2021. http://www.jstor.org/stable/41554723.

Boyd, Graham. "The Drug War Is the New Jim Crow." American Civil Liberties Union, August 2001. https://www.aclu.org/other/drug-war-new-jim-crow.

Crick, Emily. "Global Drug Policy Observatory," July 15, 2016. https://gdpo.swan.ac.uk/?p=440.

"Drug-Free Workplace Programs for Federal Agencies Background." Accessed February 10, 2023. https://www.samhsa.gov/sites/default/files/workplace/background.pdf. "Federal Sentencing Guidelines: Background, Legal Analysis, and Policy Options." EveryCRSReport.com. Congressional Research Service, March 16, 2009. https://www.everycrsreport.com/reports/RL32766.html.

Fradella, Henry F, and Michael D White. "Stop-and-Frisk." Sandra Day O'Connor College of Law. Accessed September 13, 2021.

Kernell, Samuel, and Steven S. Smith. *Principles and Practice of American Politics: Classic and Contemporary Readings*. Thousand Oaks, CA: CQ Press, 2019.

Kollman, Ken. *The American Political System*. New York: W.W. Norton & Company, Inc., 2019.

Landmark Recovery. "The History of The War on Drugs: Reagan Era and Beyond." Landmark Recovery, February 4, 2021. https://landmarkrecovery.com/history-of-the-war-on-drugs reagan-beyond/.

O'Malley, Pat, and Stephen Mugford. "The Demand for Intoxicating Commodities: Implications for the "War on Drugs"." *Social Justice* 18, no. 4 (46) (1991): 49-75. Accessed May 17, 2021. http://www.jstor.org/stable/29766642.

Oliver, Brittany. "The Racial Undertones of the War on Drugs in America." Nexus Recovery Services | Los Angeles Outpatient Treatment, February 4, 2021. https://nexusrecoveryservices.com/blog/war-on-drugs/.

Pearl, Betsy. "Ending the War on Drugs: By the Numbers." Center for American Progress, June 27, 2018. https://www.americanprogress.org/issues/criminal justice/reports/2018/06/27/452819/ending-war-drugs-numbers/.

"Race and the Drug War." Drug Policy Alliance. Accessed April 21, 2021. https://drugpolicy.org/issues/race-and-drug-war.

"Race and The War on Drugs." PBS. Public Broadcasting Service, 2019.

http://www.pbs.org/black-culture/connect/talk-back/war-on-drugs/?scrlybrkr=e ebdfc0a. Rothwell, Jonathan. "How the War on Drugs Damages Black Social Mobility." Brookings. Brookings, July 29, 2016. https://www.brookings.edu/blog/social-mobility memos/2014/09/30/how-the-war-on-drugs-damages-black-social-mobility/. "The Impact of the War on Drugs on US Incarceration ." The "militarization" of the ANTI-

DRUG Effort. Accessed September 14, 2021. https://ndsn.org/july97/military.html. Williams, Jessica. "The War on Drugs Has Damaged Black Communities and Deepened Racial Bias - IRETA: Institute for Research, Education & Training in Addictions." IRETA, June 3, 2020. https://ireta.org/resources/the-war-on-drugs-has-damaged-black communities -and-deepened-racial-bias/.

Nunn, Kenneth B. The Drug War as Race War, University of Dayton,

academic.udayton.edu/race/03justice/crime09.htm

Cooper, Hannah L F. "War on Drugs Policing and Police Brutality," March 21, 2016.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4800748/

Trine, Bill. "The Genesis of Increasing Incidents of Police Brutality: The War on Drugs." Prison Legal News. Human Rights Defense center, June 3, 2015.

https://www.prisonlegalnews.org/news/2015/jun/3/genesis-increasing-incidents-police brutality-war-drugs/

Small, Deborah. "The War on Drugs Is a War on Racial Justice." *Social Research* 68, no. 3 (2001): 896-903. Accessed May 17, 2021. http://www.jstor.org/stable/40971924. "S.1762 - Comprehensive Crime Control Act of 1984." Accessed February 10, 2023. https://www.congress.gov/bill/98th-congress/senate-bill/1762.

Zunkel, Erica, and Alison Siegler. "The Federal Judiciary's Role in Drug Law Reform in an Era of Congressional Dysfunction." SSRN, May 29, 2020.

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3589862.