



Journal of Undergraduate Research





JUR PRESS

JUR Press
Office for Undergraduate Research and Artistry
The Institute for Learning and Teaching
Colorado State University
Campus Delivery 1052
200 W. Lake Street
Fort Collins, CO, 80521-4593
Phone: (970) 491-4156
Web: jurpress.org

Designers: Ryan Rykhus
Cover Art: Fractal Trees

Copyright ©2017 JUR Press. All rights reserved.

Reproduction or translation of any part of this work beyond that permitted by Section 107 or 108 of the United States Copyright Act without the permission of the copyright owner is unlawful. The copyright of each article is held by the author. Requests for permission or further information should be addressed to the Operations Department of JUR.

ISSN: 2156-5309

Printed in the United States of America

Advertising information: If you are interested in advertising or any other commercial opportunities, please contact the Director of Operations for JUR at jurpress@gmail.com.

Author Inquiries: For inquiries relating to the submission of articles please visit jurpress.org.



EDITORIAL STAFF

Editor in Chief
Ryan Rykhus

Managing Editor
Katie Brown

Associate Editors
Alex Frickenstein
Sophie Gullett
Sydney Hardacre
Alex Keenan
Logan McCoy
Ely Roberts
Emily Ward

Copy Editors
Shannon Clare
Reagan Fitzke

Affiliate Editors

Renato Santos, Universidade Estadual Paulista, Brazil
Andreea Silaghi, Babes-Bolyai University, Romania
Nikunj Sharma, Amity University, India

OPERATIONS STAFF

Director of Operations
Graham Opie

Operations Associates
Elizabeth Cambell
Nicholas Pepping
Meryl Saxton
Maxwell You

ADVISORY BOARD

Faculty Advisor

Dr. Mark Brown
Veterinary Medicine and Biomedical Sciences
Director of the Office for Undergraduate
Research and Artistry

Industry Liaison and Contributing Editor

Jessica Egner

Faculty Advisory Board

Dr. Susan Athey
Business
Dr. Ken Blehm
Veterinary Medicine and Biomedical Sciences
Dr. Stephan Weiler
Liberal Arts
Dr. Chris Myrick
Natural Resources
Dr. Matt Hickey
Health and Human Sciences
Dr. Nancy Irlbeck
Agricultural Sciences
Dr. Don Mykles
Natural Sciences
Dr. Tom Siller
Engineering
Dr. Melissa Edwards
University Honors Program
Dr. Rita Zapata Vázquez
Universidad Autónoma de Yucatán
Dr. Kiley Miller
Schreiner University



A Letter from the Editor in Chief

JUR Press is proud to present Volume VIII of the Journal of Undergraduate Research with published works across a wide range of disciplines by undergraduates from around the world.

We are a journal for undergraduates and by undergraduates. Our slogan – linking the global undergraduate community – captures our mission to serve the interests of worldwide undergraduate thinkers, tinkerers, experimenters, writers, and artists. While JUR does provide a platform for students to publish and showcase their work, we also give students the opportunity to learn about the publication process from start to finish through internships as editors, operations associates, and reviewers. Our undergraduate network has grown to include not only international authors, but also affiliate and satellite editors from around the world! We are truly committed to enhancing the undergraduate experience, and we continually seek to engage as many undergraduate students in our organization as possible.

This year has been a blessing as I have settled into the role of editor in chief with support from those around me. I am honored to have such a strong foundation built from all staff members and the entire JUR network. As seniors graduate and pursue their future and new members join to engage in the publication process, I am reminded of the hard work, determination, and dedication every present and past JUR member has contributed to better the organization and themselves. I have seen each member this year grow and truly embrace the meaning of “interdisciplinary” as they manage academics, internships, sports, volunteer positions, and take time out of their day to make someone else’s. I am incredibly thankful for the laughs and smiles that each unique personality brought throughout the year to complete this issue.

For all that we have accomplished this year, I want to say congratulations to each member and how incredibly thankful I am for their efforts. You are exceptional individuals with a wide range of talents. I look forward to what the future has to offer for those graduating, and what the year holds for new and returning members.

To our published authors, I congratulate each of you on your accomplishment and we are thankful to have been part of your undergraduate academic experience.

Yours truly,

Ryan Rykhus
Editor in Chief
Journal of Undergraduate Research
JUR-Press

Special Thanks

This publication would not have been possible without Colorado State University and the contributions of numerous advocates and benefactors.

JUR Press would like to thank our Faculty Supervisor, Dr. Mark Brown, the Faculty Advisory Board, and our Graduate Advisor Melissa Edwards. We would also like to thank our partners at the Autonomous University of the Yucatán, in Mexico, and Schreiner University in Texas. Finally, thanks to Jessica Egner, JUR’s first editor in chief, for her past and ongoing support as Industry Liaison and Contributing Editor.

Congratulations to our graduating seniors: Sophie Gullett, Graham Opie, and Emily Ward



TABLE OF CONTENTS

TITLE	AUTHOR	PAGE
In a cinch	By Jaelyn Larson <i>Colorado State University</i>	6
Into the Green	By Emily Michaud <i>Colorado State University</i>	7
Midnight Festoon	By Melissa Leroux <i>Colorado State University</i>	8
Fractal Trees	By Joshua Thomas <i>Colorado State University</i>	9
The Biosynthesis and Synthetic Biology of Artemisinin	By Joseph R. Stephenson Clarke <i>University of Reading</i>	10
Lactate Promotes Survival of Peripheral Blood Mononuclear Cells by Decreasing Apoptotic Events	By Sarah L.C. Elsaser, Fiona F. Vickers, Catalina M. Carvajal Gonczi, Peter J. Darlington <i>Concordia University, Montreal, Quebec, Canada</i>	17
Whole Exome Sequencing and Analysis to Rule Out Three Bone-Expressed Genes as Candidates for Split Hand Foot Malformation	By C. Ronnie Funk, Melanie M. May, Anna V. Blenda, Ph.D., & Charles E. Schwartz, Ph.D. <i>Erskine College</i>	21
Exercise and Autism Spectrum Disorder: A Neuroscience Perspective	By Alex P. Kunz, Mentor Brock Jensen, PhD <i>Slippery Rock University</i>	26
The Role of Impulsivity in the Effectiveness of Restrictive Eating Behaviors	By Sarah Nachimson <i>Stern College for Women, Yeshiva University</i>	30
Female War Correspondents and their Effect on Female Military Positions	By Karlin Anderson <i>Whitworth University</i>	32
The Royal Navy's Blockade System 1793-1805: A Tactical Paradox	By Patrick Braszak <i>Wilfrid Laurier University</i>	35
A Song of Burning Wood: Colonization and Memory in A Song of Ice and Fire	By Hannah Gildart <i>Southwestern University</i>	41
And There They Were: Mrs. Dalloway, British Identity and the Indian Other	By Joelle Hamilton <i>Colorado State University</i>	44
Picking the Pretty One: A Qualitative Study of Toy Selection Among Girls Age Four to Ten	By Savannah Anderson-Bledsoe and Donna S. Sheperis, PhD. <i>Lamar University</i>	48
Agrarian Capitalism and the Privatization of Collective Land in Morocco	By David Balgley <i>University of Puget Sound</i>	55
Work-Related Soft Skills Summer Program: A Case Study	By J. Christian Banez & Shayda Afrassiab <i>Minnesota State University-Mankato</i>	63
Culture as Resistance: A Study of the Warsaw and Łódź Ghettos	By Miranda Brethour <i>University of Ottawa</i>	67
Troubling the Nation: Crafting the New Gay in a Post 9/11 Environment	By Tara A. Casey, Martha J. Garner, Rebekah S. Kamp, & Gabriel Villarreal <i>Northern Arizona University</i>	74
Conceptualizing Trauma by Embracing Ambiguity	By Audra DeBoy <i>Gettysburg College</i>	78
Domestic Violence and Women's Empowerment in Nepal	By Zhicheng Han <i>Brigham Young University</i>	83



ART



In a cinch By Jaelyn Larson

For this project, I desired to study different ways a garment could be hung, draped, or manipulated to create a visually intriguing composition. I chose my favorite coat because its cinched drawstring allowed for complex folds and wrinkles. I hope this piece resonates with artists and fashion lovers alike!





Into the Green
By Emily Michaud

This work focuses on the illustrating of initial impressions based upon descriptive imagery and ideas found within the pages of books. In this piece, I investigate the way that multiple perspectives interact with one another, as well as the often dynamic relationship between natural and man-made structures. I work to create interesting, thought provoking illustrations through the use of whimsy, watercolor, and other various forms of media.





ART



Midnight Festoon
By Melissa Leroux





Fractal Trees
By Joshua Thomas



The Biosynthesis and Synthetic Biology of Artemisinin

BY JOSEPH R. STEPHENSON CLARKE
UNIVERSITY OF READING

Abstract

Artemisinin is a frontline antimalarial drug that occurs naturally in the native Chinese herb Artemisia annua. This natural product has a complex structure featuring an unusual 1,2,4-trioxane ring which is responsible for its powerful antimalarial activity. The natural biosynthesis and recent advances in the synthetic biology of artemisinin are discussed in this review. The biosynthesis is divided into three phases. The first involves the head-to-tail linkage of five-carbon allylic pyrophosphates to form a longer (fifteen-carbon) chain pyrophosphate intermediate, which is then cyclized to the bicyclic amorphane skeleton of artemisinin, amorpha-4,11-diene (A-4,11-D). In phase two A-4,11-D undergoes both oxidation and reduction to dihydroartemisinic acid (DHAA). The mechanism of the third and final phase is still subject to some controversy, though it must involve oxidative cleavage of the C-4 double bond in one of the rings of DHAA. It is also possible that this final step is a spontaneous autoxidation and that the 1,2,4-trioxane ring of artemisinin is generated non-enzymatically. Over the past decade there has been much interest in synthetic biological approaches to the production of artemisinin, which employ many of the enzymes that have been described from the biosynthetic pathway to artemisinin in A. annua. In the industrial process, as it is currently practiced, the genes that code for these enzymes are expressed in the chassis organism Saccharomyces cerevisiae (yeast). This produces artemisinic acid (AA), a close relative of DHAA, which is then converted to artemisinin by chemical synthesis.

Introduction

Malaria is an aggressive and highly infectious parasitic disease that is prevalent throughout more than 100 equatorial and subtropical countries. In 2006, 86% of documented cases of malaria occurred in Africa; over half of these were reported in Nigeria, Ethiopia, the Democratic Republic of the Congo, Kenya, and the United Republic of Tanzania.¹ This figure was reported by the World Health Organization (WHO) to have increased to 90% in 2016. Since malaria is largely a Third-World disease, a cheap and effective cure is in great demand.

Due to rising concerns about chloroquine-resistant strains of the malarial parasite *Plasmodium falciparum* during the Vietnam War, a project to screen traditional Chinese medicines (TCM) for antimalarial agents was commissioned by the Chinese government² (historically this was parallel to the U.S. discovery of mefloquine).³ The potent antimalarial activity of *A. annua*, for which usage in TCM dates as far back as 168 B.C.E., was recognized by a team led by the pharmaceutical chemist Youyou Tu. Professor Tu was subsequently awarded the 2015 Nobel Prize for Physiology or Medicine for her contribution to the development of artemisinin. The Chinese scientists characterized artemisinin as a sesquiterpene (fifteen-carbon) lactone containing an endoperoxide. It was noted that this endoperoxide group was essential to

its antimalarial function. In clinical trials artemisinin demonstrated 100% effectiveness towards chloroquine-resistant strains of the *P. falciparum* parasite.⁴

In its natural setting artemisinin probably functions as a chemical defense for *A. annua* against attack by herbivores and insects.^{5,6} It is quite common for natural products with such an “ecological” role to be useful in the treatment of human diseases; in addition to its powerful antimalarial function, artemisinin also shows antiviral and anticancer activity.⁷

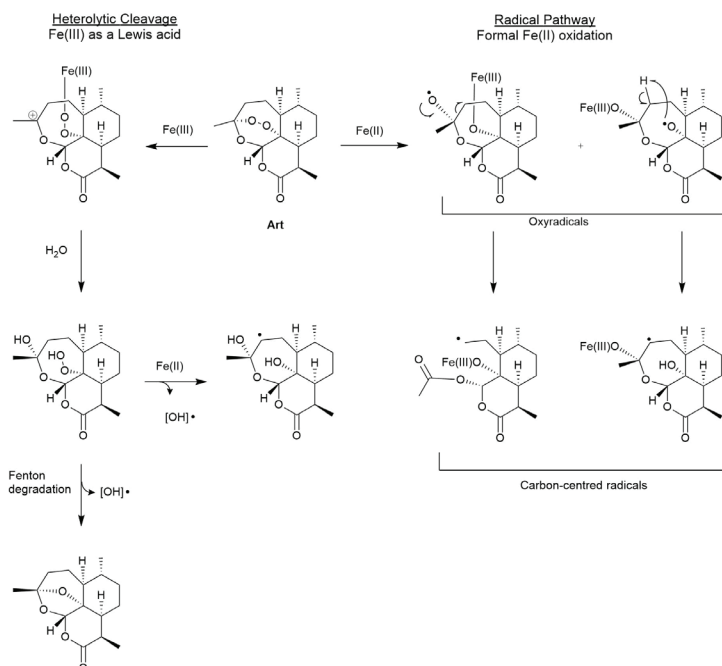
Two main pathways have been proposed for the mode of action of artemisinin (Art). In both, the endoperoxide group is involved in the *in vivo* activation of the drug *via* interaction with iron(II) atoms. These may come from free heme groups released by the digestion of hemoglobin by the parasite, although there is evidence for bio-activation by both heme iron and non-heme iron(III).⁸ In one pathway the endoperoxide group of Art is heterolytically cleaved in a reaction facilitated by the Lewis acidity of iron, followed by interaction with water, which generates a hydroperoxide capable of directly oxidizing protein residues.⁸ This hydroperoxide can also undergo Fenton degradation, thereby expelling a hydroxyl radical (Scheme 1). In the second radical pathway iron(II) is oxidized by either of the endoperoxide oxygen atoms, generating an oxyradical which re-

arranges to a carbon-centered radical capable of both heme- and protein-alkylation (Scheme 1). Some of the proposed carbon-centered radical intermediates have been detected by EPR and spin-trapping studies.⁹

An alternative model involves depolarization of the parasite’s mitochondrial and plasma membranes *via* lipid peroxidation by a reactive oxygen species that is also generated from iron bio-activation of the endoperoxide.^{8,10} There is experimental evidence that physiological levels of artemisinin can depolarize parasite plasma ($\Delta\Psi_p$) and mitochondrial ($\Delta\Psi_m$) membrane potentials.⁸

In practice it is normal to use chemically-modified derivatives of artemisinin in place of the parent compound due to its poor solubility and unfavorable metabolic and hydrolytic stabilities. Generally such an artemisinin derivative is administered alongside another antimalarial in a treatment known as artemisinin-based combination therapy (ACT). ACT is advantageous in that treatment times are reduced and the evolution of drug-resistant strains of the malarial parasite is hindered. ACTs are now recommended by the WHO as the treatment of choice for *falciparum* malaria.

Since its discovery the main problem with this drug has been that the worldwide market for artemisinin is highly unstable due to its dependence on agriculturally-produced artemisinin from an annual crop



Scheme 1: Heterolytic and radical cleavage mechanisms proposed for the iron-catalyzed bio-activation of artemisinin in vivo.⁸

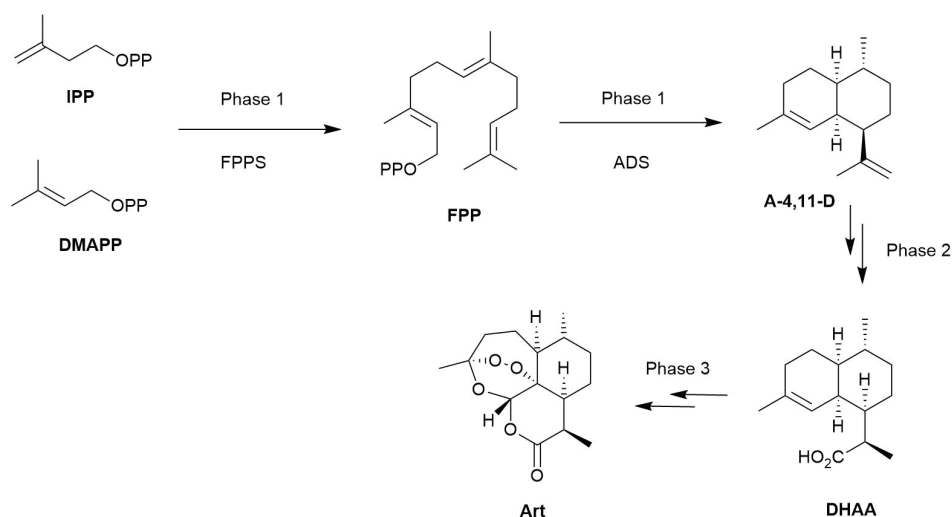
cycle. However, in recent years, a promising development towards a stable supply of the antimalarial drug has come from the emerging science of synthetic biology, in which artemisinin precursors are produced from simpler single-celled organisms in a fermentation process. As discussed below the effectiveness of synthetic biology as an alternative to the agricultural production of artemisinin is intimately linked to our understanding of its biosynthesis in *A. annua*, the natural source.

Biosynthesis

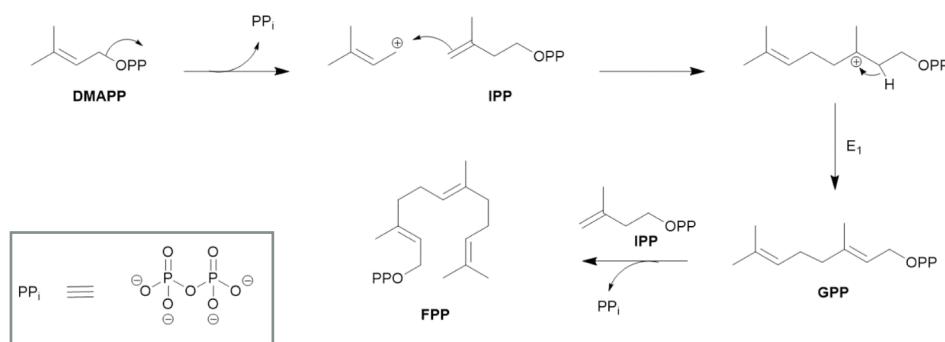
The biosynthesis of artemisinin (Art) in *A. annua* is typically considered in three phases, as shown in Scheme 2.

Phase 1: Biosynthesis of amorpha-4,11-diene (A-4,11-D)

Phase one of the biosynthesis was established first and is the best understood of the three phases. Most of the enzymes involved in phase one have been isolated and fully characterized directly from *A. annua*, providing evidence for their involvement in the proposed biosynthetic pathway. Artemisinin is sequestered in biserial glandular trichomes on the surface of the leaves of *A. annua* since it is highly autotoxic to the plant. Substantial evidence has now been presented to suggest that these enzymes are also located in the trichomes, which are the sites in which the majority of the biosynthesis takes place.^{5,6,11}



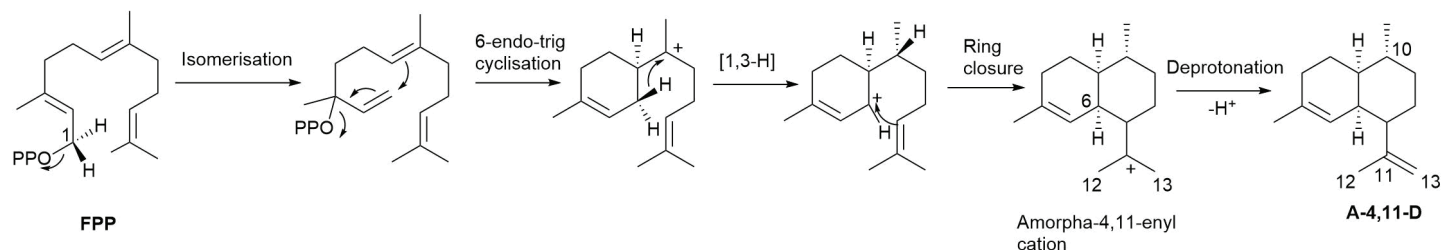
Scheme 2: The three phases in the biosynthesis of artemisinin.



Scheme 3: Stepwise chain extension of DMAPP to FPP by the enzyme FPPS.

Carbon-14 labelling experiments have proven that the hemiterpenoid (five-carbon) isopentenyl pyrophosphate (IPP) is one of the first biogenic precursors to artemisinin.¹² The enzyme farnesyl pyrophosphate synthase (FPPS) catalyzes the successive head-to-tail linkage of two units of IPP with its isomer, dimethylallyl pyrophosphate (DMAPP), in order to generate farnesyl pyrophosphate (FPP) via the intermediate geranyl pyrophosphate (GPP) (Scheme 3). This pseudo-oligomerization process is driven by the stability of the pyrophosphate (PP) leaving group. When the stable pyrophosphate moiety is expelled from DMAPP, an allylic carbocation is formed which is subsequently attacked by the double bond of IPP. This is followed by an E₁ elimination step which reforms the alkene. The same mechanism is then repeated with GPP as the substrate and a second unit of IPP. Both the hemiterpenoid precursors IPP and DMAPP are biosynthesized predominantly via the non-mevalonate (DXP) pathway in *A. annua*,³ although there is evidence for some cross-talk with the alternative mevalonate pathway.¹³

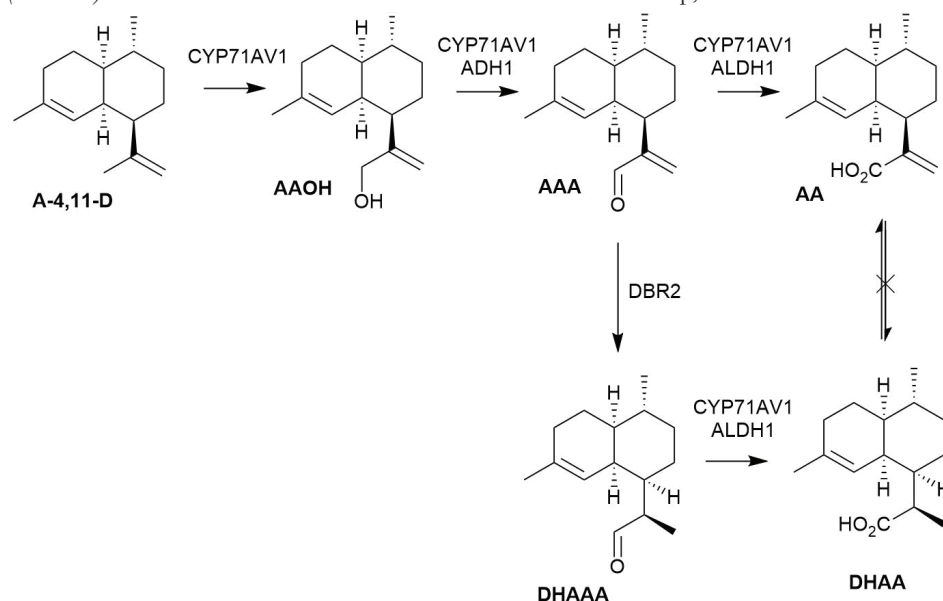
Next, FPP is cyclized to key intermedi-



Scheme 4: Mechanism of the ADS-catalyzed cyclisation of FPP to amorpha-4,11-diene (A-4,11-D).

ate amorpha-4,11-diene (A-4,11-D) by the enzyme amorphadiene synthase (ADS). Like FPPS, ADS is most highly expressed in the glandular secretory trichomes of *A. annua*.¹⁰ The mechanism of ADS catalysis was first demonstrated by the Kim group in 2006 from experiments using deuterium-labeled FPP. They showed that following isomerization of the pyrophosphate group in FPP, this group is lost in a concerted 6-endo-trig cyclization which generates a tertiary carbocation. H-1 of FPP then migrates to the 10-position of amorpha-4,11-diene (Scheme 4). It was possible to identify which of the two diastereotopic C-1 hydrogens underwent this 1,3-hydride shift [1,3-H] by using singly-labelled FPP.¹⁴ Thus, experiments using both enantiomers of singly-labeled [1-²H] FPP confirmed that while one hydrogen migrates to the 10-position of A-4,11-D, the other remains in its place as the H-6 of the amorpha-4,11-enyl cation.¹⁴ The final deprotonation step can occur either from C-12 or C-13 of the amorpha-4,11-enyl cation, resulting in the product, amorpha-4,11-diene (A-4,11-D).³

Phase 2: Biosynthesis of dihydroartemisinic acid (DHAA)



Scheme 5: Key intermediates in phase 2 of the conversion of A-4,11-D to AA or DHAA.

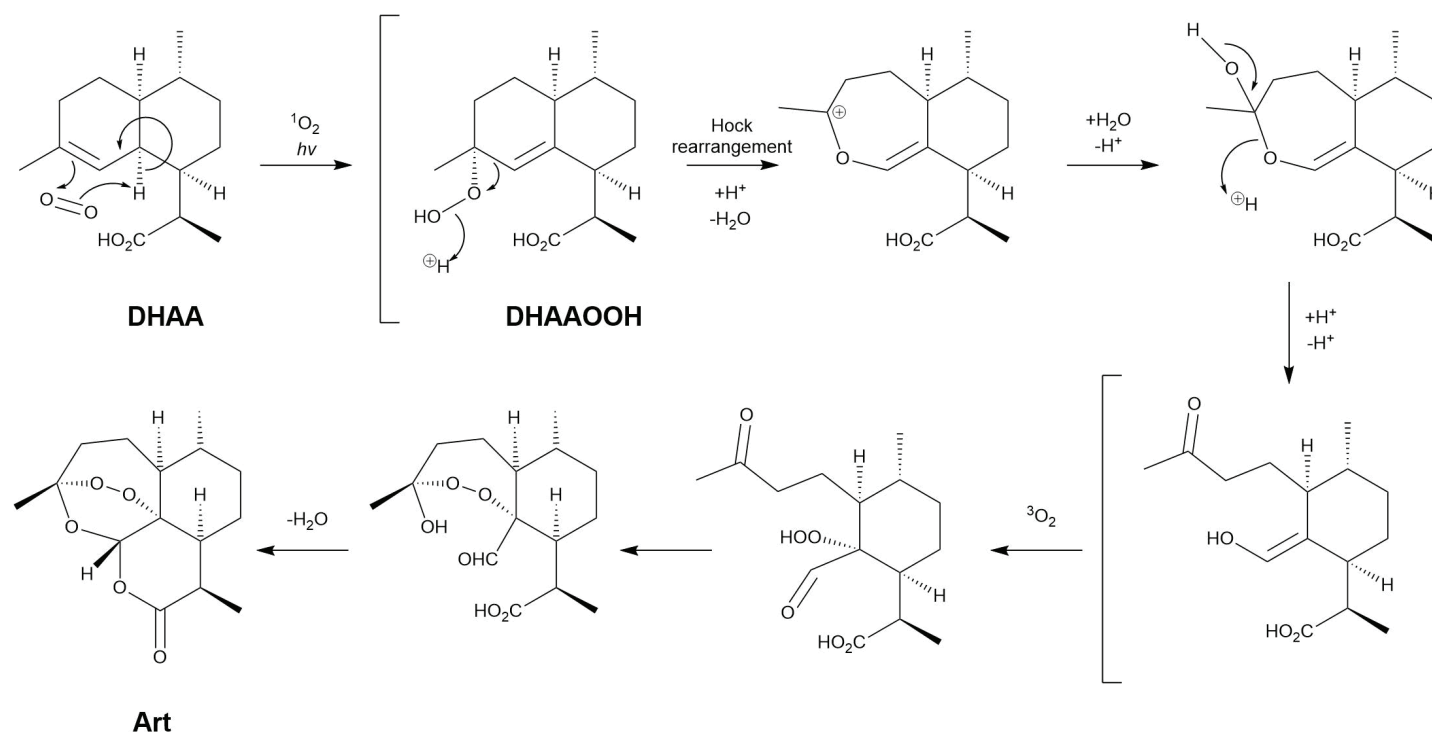
More recently it has become well-established that the oxidative transformation of A-4,11-D yields two alternative products — artemisinic acid (AA) and dihydroartemisinic acid (DHAA). Several of the steps that potentially feature in the conversion of amorpha-4,11-diene to AA and DHAA are inferable from natural products which have been isolated directly from the plant. Thus, the characterization of artemisinic alcohol (AAOH) and artemisinic aldehyde (AAA) from *A. annua* are both consistent with a biosynthetic pathway that involves the oxidation of the isopropylidene functionality in A-4,11-D, as shown in Scheme 5.

For many years AA was assumed to be the biogenic precursor to artemisinin in phase 3, though it is now thought to be a by-product of the biosynthesis, with DHAA as the true precursor. Significant evidence has now accumulated for the route to DHAA that is shown in Scheme 5, which involves two sequential oxidations of A-4,11-D at the C-12 position to AAA, followed by a reduction of the 11,13-double bond to dihydroartemisinic aldehyde (DHAAA). DHAAA is then oxidized once again at C-12 to form DHAA. The gene DBR2, which catalyzes the reduction step, has been demonstrated

to be highly expressed in the glandular secretory trichomes, which are the likely site of artemisinin production in *A. annua*.^{5,6,11} Recombinant DBR2 has also been demonstrated to be relatively specific for the reduction of AAA — for example, it is not capable of reducing the 11,13-double bond in some closely related natural products from *A. annua*, such as AA, AAOH, arteannuin B, and artemisitene.^{13,15} Although DBR2 exhibits activity towards a few small α,β -unsaturated carbonyl compounds (as would be expected for a member of the enoate reductase family of enzymes), it was shown to be unreactive to several others including coniferyl aldehyde, 2*E*-nonenal, and 12-oxophytodienoic acid.^{11,13} Similarly a cDNA for the cytochrome P450 enzyme CYP71AV1 has been shown to be capable of effecting all three of the oxidations of A-4,11-D featured in Scheme 5, in combination with a P450 reductase, and both have been demonstrated to increase artemisinin content in transgenic *A. annua* plants significantly.¹³ Most recently a cDNA encoding the aldehyde dehydrogenase homolog ALDH1, which is capable of facilitating the final oxidation of DHAAA, has also been characterized and demonstrated to be highly expressed in the trichomes.¹³

One of the reasons that AA has been discounted as a biogenic precursor to artemisinin over recent years has been the observation that there is little correlation between AA and artemisinin contents when comparing different transgenic *A. annua* strains.¹⁶ In contrast, levels of artemisinin are found to be positively correlated to levels of DHAA across a number of genotypes of *A. annua* (*i.e.* the more DHAA that is observed in a strain, the more artemisinin that strain is capable of producing).¹⁶ In addition there is also a direct negative correlation between DHAA and artemisinin content *within* any given strain (*i.e.* the less DHAA that is observed in any given specimen, the more artemisinin is present). This negative correlation is most readily rationalized by the assumption that DHAA can be converted to artemisinin and is one piece of evidence that DHAA is the biogenic precursor to artemisinin in the next phase of the biosynthesis.

Note, however, that AA and DHAA



Scheme 6: A possible mechanism for the spontaneous autoxidation of dihydroartemisinic acid to artemisinin.

contents are positively correlated to one another, as would be expected from their relationship in Scheme 5.

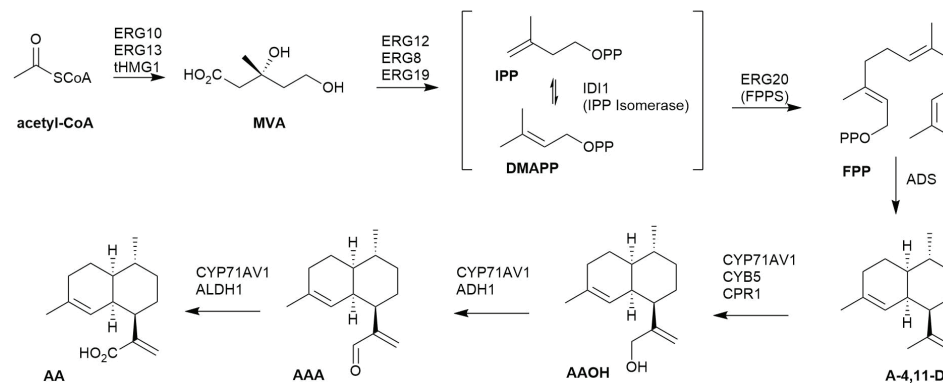
Phase 3: Biosynthesis of artemisinin (Art)

The third and final phase of artemisinin biosynthesis is the least understood, although there is some evidence to suggest that the third phase can proceed non-enzymatically in the plant.¹⁷ Under experimental conditions mimetic of chloroplasts, Wallaart *et al.* have proven that singlet oxygen ($^1\text{O}_2$) enhances yields of artemisinin, although chlorophyll is not necessary for the conversion, which is thought to take place in the trichomes.^{17,18} The proposed pathway involves photooxidation of DHAA by an ene-type mechanism to dihydroartemisinic acid hydroperoxide (DHAAOOH) (Scheme 6), which was identified in Wallaart's experiments, and is a known natural product from *A. annua*.^{18,19} Hock rearrangement of the allylic hydroperoxide generates an oxacyclic intermediate that is ring-opened to an enol.^{15,18} This then undergoes a second oxygenation by triplet oxygen ($^3\text{O}_2$) to generate a hydroperoxyl-aldehyde, which is cyclized to the 1,2,4-trioxane system. The comparative ease by which DHAA can be converted to artemisinin *via* DHAAOOH *in vitro* provides circumstantial evidence for a non-enzymatic route for the third phase of natural production.¹⁹

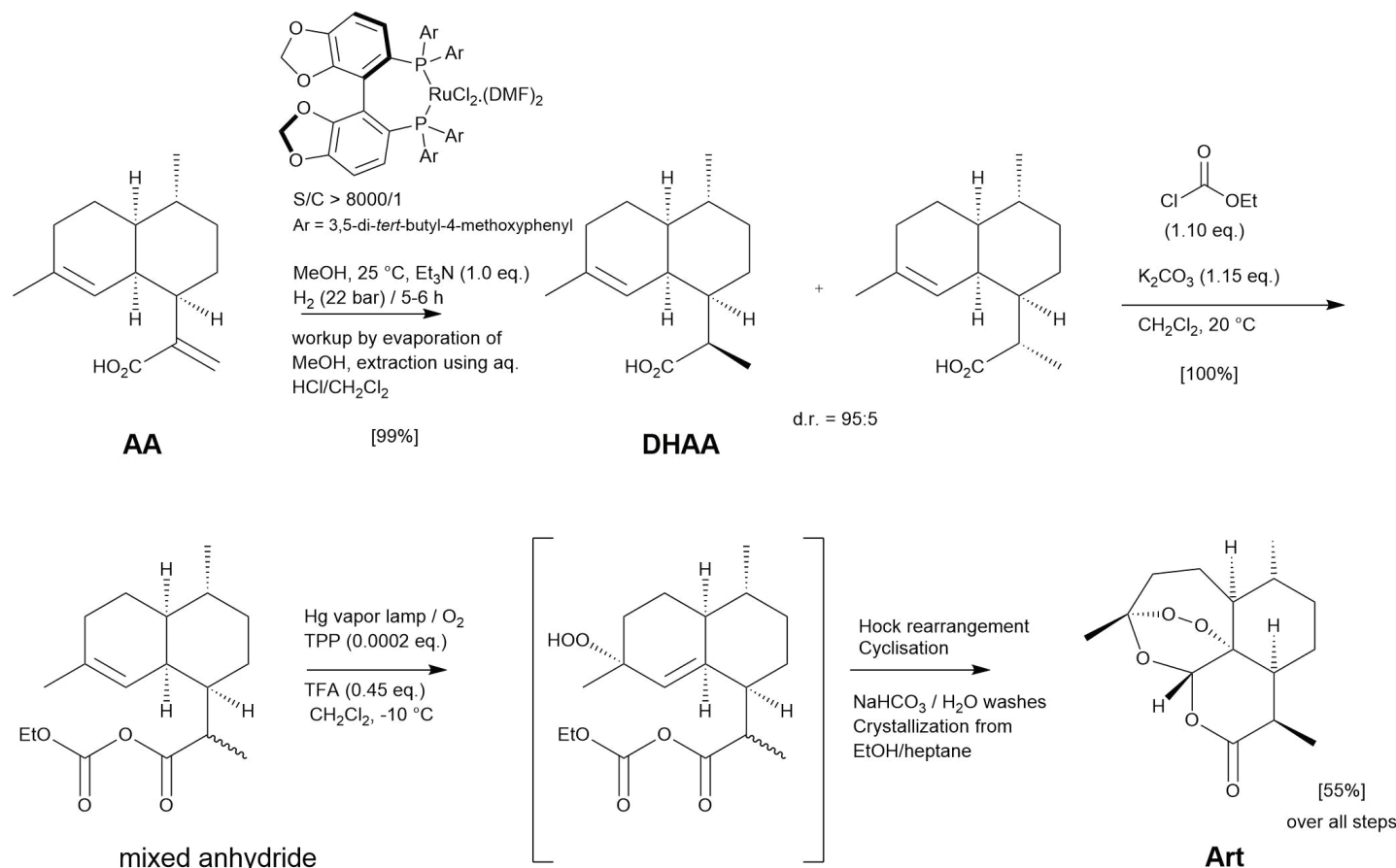
Synthetic biology

Synthetic biology is the genetic engineering of simple organisms to construct useful and exploitable biological systems. Recombinant DNA programming can be used to manipulate metabolic networks in order to transform a microbial host, such as *Saccharomyces cerevisiae* (yeast) or *Escherichia coli* (a bacterium), into a cheap "chemical factory" for the production of natural products.²⁰ Natural products are good targets for synthetic biology since their biosynthetic pathways have often been studied and documented in detail. Additionally, many natural products are expensive and available only in limited quantities from the natural source.

The mean concentration of artemisinin in dry *A. annua* leaves is typically less than 0.5%,²¹ although Swiss-developed transgenic strains have reached almost 2% dry-leaf artemisinin content.²² Direct metabolic engineering of the plant is therefore demonstrably feasible and, as such, offers the most convenient means for increasing yields of agriculturally-produced artemisinin, as well as providing potential routes to novel alternative sesquiterpenes.^{17,22} Synthetic biology, by contrast, holds out the promise of much higher yields of artemisinin, even though it is a technically more demanding approach that requires not only the transfer of relevant genes from the natural source to the micro-



Scheme 7: Conversion of acetyl-CoA to AA in *S. cerevisiae* - the genetically-engineered component of the industrial semi-synthesis of artemisinin



Scheme 8: Chemical transformations of AA to artemisinin (Art) in the chemical synthesis component of Sanofi's artemisinin semi-synthetic process.²⁹

bial host, but also the enhancement of gene expression in that host by the editing of gene promoters and terminators.^{23,24}

Promoters are sequences in DNA which influence the transcription of a gene by providing a binding site for RNA polymerase. Promoters thereby regulate the production of mRNA, thus controlling protein levels inside a cell.²⁴ Useful promoters are typically obtained by constructing libraries of synthetic promoters that cover a range of incrementally increasing activities, thereby permitting quantitative tuning of gene expression in bioengineering.²⁵

In 2003 Prof. Jay Keasling published the first landmark semi-synthesis of artemisinin using genetically engineered organisms (the term "semi-synthesis" emphasizes that his procedure incorporated elements of both synthetic biology and chemical synthesis). Capitalizing on the success of his small team at the University of California, Berkeley, Keasling then co-founded the biotechnology firm Amyris Inc. and began "The Semi-Synthetic Artemisinin Project," with the intent of delivering a new and cheap alternative source of the drug to stabilize worldwide

demand. The Project involved the bio-engineering of the microbial host *S. cerevisiae* to produce artemisinic acid (AA). *S. cerevisiae* was preferred as the host for the microbial biosynthesis of AA since the yield was greater than in *E. coli* (40 g/L as compared with 25 g/L) – ultimately, this was because the latter suffered from issues in the oxidation of A-4,11-D.²⁰

After mechanical recovery of AA from the yeast's cell wall, it was converted to artemisinin by chemical synthesis. In 2008 Amyris granted the recipe for their semi-synthetic procedures royalty-free to French pharmaceutical company Sanofi, in order to set up a large-scale production facility. In 2014, Sanofi was able to distribute over 1.7 million doses of microbially-derived ASAQ (an ACT) to six African countries, including Niger, Nigeria and Liberia.²⁶

The genetically-engineered component of the artemisinin semi-synthesis

In the industrial process it is the mevalonate pathway that is naturally present in yeast that provides the sole source of IPP and DMAPP. This pathway is fundamentally

different from the DXP pathway in *A. annua* and is named after the six-carbon intermediate (R)-mevalonic acid (MVA), which is biosynthesized from acetyl-CoA by three enzymes (ERG10, ERG3, tHMG1), and is then decarboxylated to the hemiterpenes IPP and DMAPP by ERG12, ERG8 and ERG19 (Scheme 7). All these ERG genes (as well as IDI1 and tHMG1 in Scheme 7), code for enzymes in the ergosterol biosynthetic pathway (ergosterol is an essential steroidal component of fungal membranes).²⁰ It is also possible to transfer the genes encoding the "natural" *A. annua* DXP pathway for isoprenoid synthases – indeed, these have been successfully over-expressed in *E. coli* in order to accelerate isoprenoid synthesis in this microbial host.

Next, FPP is bio-synthesized from IPP and DMAPP by ERG20 (farnesyl pyrophosphate synthase, catalyzing the same reaction as in *A. annua* — see Schemes 2 and 3). All the genes used in the industrial process in yeast subsequent to FPP synthesis are derived from *A. annua*. Thus FPP is cyclized to amorphaadiene (A-4,11-D) by ADS (see Schemes 2 and 4) and A-4,11-D undergoes

three successive oxidations by CYP71AV1 derived from *A. annua* (as in Scheme 5). These oxidations are facilitated by CYB5 and CPR1, which significantly boost the yield of microbial artemisinic acid (AA),²⁷ as in the plant. Amyris have also reported that the over-expression of genes encoding ERG20 biosynthesis doubles AA yields in *S. cerevisiae*.²⁸

The chemical synthesis component of the artemisinin semi-synthesis

The chemical processing that is required for the conversion of microbial AA to artemisinin involves two steps: an initial reduction of the 11,13-double bond in AA to form DHAA, followed by introduction of the 1,2,4-trioxane ring. In the industrial process the challenging diastereoselective hydrogenation of AA to DHAA has been optimized using a chiral ruthenium catalyst under H₂ at a pressure of 22 bar, to afford 95% selectivity for the natural product (Scheme 8). The photo-catalyzed second step proceeds by the same complex mechanism that is thought to occur non-enzymatically in *A. annua*, as is shown in Scheme 6. However the industrial oxidation reaction is performed in dichloromethane using tetraphenylporphyrin (TPP) as a sensitizer for the photochemical conversion of ³O₂ to ¹O₂, rather than in the lipophilic environment of a trichome. In addition the natural intermediate DHAAOOH was found to be too unstable for the industrial one-pot process. It was discovered that prior derivatization of DHAA to a mixed anhydride facilitated a cleaner conversion to the requisite transient allylic hydroperoxide, with no hazardous accumulation of this intermediate in the complex reaction cascade required for the subsequent oxidation that forms the 1,2,4-trioxane ring (Scheme 8).²⁹

It is worth noting that even though the current method of chemical synthesis affords a 55% overall yield (which represents a considerable improvement on 19% for the original synthetic conversion of AA to artemisinin),²⁹ the complicated chemical steps that are shown in Scheme 8 still account for most of the production costs of semi-synthetic artemisinin.³⁰ As a result optimization of the chemical steps of the Sanofi semi-synthesis is the most studied approach to improving industrial artemisinin yields.

Current approaches to improving industrial artemisinin production

Recently a greener method using liquid CO₂ solvent has been presented as an alternative to the above photooxidation — however, it is lower yielding and thus more expensive.^{30,31} It could be argued that, in the case of artemisinin, decreasing the cost of treatments is more important than environ-

mental considerations.

Despite the expected safety issues Sanofi developed a diimide-based reduction as a metal-free alternative to the catalytic hydrogenation shown in Scheme 8. This approach used the reaction of oxygen and hydrazine monohydrate to generate diimide *in situ*, which effected the desired reduction of AA to DHAA in >90% yield with excellent diastereoselectivity (d.r. ≥ 97:3).³² This method was successful on the kilogram scale and presents a viable alternative to the traditional catalytic hydrogenation.³²

The Seeberger group developed a continuous-flow method as an alternative to the Sanofi photooxidation of DHAA.³³ One of the greatest limitations of large-scale photochemistry is that light can only penetrate a short distance through a solution before it is absorbed by a photosensitizer; thus the use of large vessels results in reduced conversion rates. Seeberger's flow system circumvented this limitation to photochemical generation of ¹O₂ by wrapping a tube around a lamp to maximize light absorbance. In addition to the usual benefits granted by the use of flow chemistry (such as good control over reaction parameters, easy scale-up and safety) Seeberger's method allowed the conversion of DHAA to artemisinin in a continuous process requiring no isolation or purification of intermediates.³³ Improvements on the original continuous-flow setup resulted in a system capable of producing 165 g of artemisinin per day.³⁴ Unfortunately this process may not be commercially viable due to the yield and the requirement for chromatographic purification of the product.³¹

Very recently a novel semi-synthesis of artemisinin was published which uses A-4,11-D as the chemical precursor.³¹ A-4,11-D is formed earlier than AA in the bioengineered pathway and in twice the molar yield;³¹ thus chemical synthesis of artemisinin from A-4,11-D presents a promising alternative to the currently practiced semi-synthesis. A cost-effective semi-synthetic route was devised which could produce artemisinin from A-4,11-D in 6 steps in 60% yield. It was predicted that this novel approach could reduce global artemisinin costs from 200 USD/kg (2016) to around 100 USD/kg.³¹

Conclusion

It can be argued that the development of the industrial process reflects the progress that has been made in our understanding of the three phases of the biosynthesis of artemisinin. Thus the genetically-engineered component of the synthetic biology approach to artemisinin has benefitted greatly from our discoveries regarding phases 1 and 2 of its biosynthesis. By contrast there is still much uncertainty in our understanding of

the details of phase three of the biosynthesis; this finds a parallel in the requirement for a chemical synthesis component the industrial process (in particular, the chemical step that mimics the mechanism of the spontaneous autoxidation described in the plant). As a result, although microbially-sourced artemisinin is now providing a viable alternative source of this drug that is already supporting the improving worldwide supply of ACT treatments, it is no cheaper than agriculturally-produced artemisinin. This may change in the future as improvements in our understanding of the natural biochemistry continue to inform the development of key steps in the industrial process.

One obvious prediction is that we should expect to see a continuation of research into the transfer of DBR2 and ALDH1, the two most recently discovered genes from phase 2 of the biosynthesis in *A. annua* (Scheme 5), to a microbial host. This is expected to be a fruitful line of enquiry because the successful expression of DBR2 and ALDH1 in *S. cerevisiae* could result in a genetically-engineered component of the industrial process that produces DHAA rather than AA. This should, in turn, lead to a substantial reduction in the cost of industrial artemisinin production because it would obviate the need for the first (reductive) step in the chemical synthesis component of the process. (As shown in Scheme 8, this is one of two obligatory chemical steps for the chemical conversion of artemisinic acid to artemisinin, both of which are low-yielding and expensive in comparison with the genetically-engineered component). In contrast semi-synthesis from the earlier biogenic precursor A-4,11-D could present a cost-effective alternative to the current industrial process.

Looking further into the future it may also be possible that a better understanding of phase 3 of the natural biosynthesis in *A. annua* could inspire a more fundamental revolution in the industrial production of artemisinin. If it were discovered that an enzyme exists which can catalyze phase 3 of the biosynthesis (in addition to the spontaneous autoxidation pathway that has already been described), the industrial synthesis may be developed to eliminate the need for the expensive chemical conversion of DHAA. This might be a specific protein from *A. annua* that is capable of converting DHAA to artemisinin, or, alternatively, a family of oxidases from some other species that fulfils this same function. Much effort would surely then follow in the attempt to transfer the gene(s) for this enzyme to a microbial host. If this were achieved, in combination with the transfer of genes for DBR2 and ALDH1 that is currently in progress, then it could one day become possible to express the

entire biosynthetic pathway to artemisinin in a microorganism. This would be a great triumph for synthetic biology: an industrial process that is based solely on genetic engineering - inevitably greener than the current process and certainly cheaper than agricultural methods - making artemisinin more accessible than ever.

Acknowledgements

I would like to thank Dr. Geoffrey Brown for the inspiration that led to this project and for his kind support throughout.

References

- N.A. (2008) "World malaria report 2008." *World Health Organization*, World Health Organization. <<http://apps.who.int/iris/bitstream/>> (Accessed 03/12/2016)
- Miller, L.H. and Xinzhuan, S. (2011) "Artemisinin: discovery from the chinese herbal garden." *Cell*. 146.6. Pg 855-858.
- Schramek, N., Wang, H., Römisch-Margl, W., Keil, B., Radykewicz, T., Winzenhörlin, B., Beerhues, L., Bacher, A., Röhlich, F., Gershenzon, J., Liu, B. and Eisenreich, W. (2010) "Artemisinin biosynthesis in growing plants of *Artemisia annua*. A 13CO₂ study." *Phytochemistry*. 71.2-3. Pg 179-187.
- Sriram, D., Rao, V. S., Chandrasekhar, K.V.G. and Yogeewari, P. (2004) "Progress in the research of artemisinin and its analogues as antimalarials: an update." *National Product Research*. 18.6. Pg 503-527.
- Duke, M.V., Paul, R.N., Elsohly, H.N., Sturtz, G. and Duke, S.O. (1994) "Localization of artemisinin and artemisitene in foliar tissues of glanded and glandless biotypes of *Artemisia annua* L." *International Journal of Plant Sciences*. 155.3. Pg 365-372.
- Ferreira, J.F.S. and Janick, J. (1995) "Floral morphology of *Artemisia annua* with special reference to trichomes." *International Journal of Plant Sciences*. 156.6. Pg 807-815.
- Barbacka, K., and Baer-Dubowska, W. (2011) "Searching for artemisinin production improvement in plants and microorganisms." *Current Pharmaceutical Biotechnology*. 12.11. Pg 1743-1751.
- Antoine, T., Fisher, N., Amewu, R., O'Neill, P.M., Ward, S.A. and Biagani, G.A. (2014) "Rapid kill of malaria parasites by artemisinin and semi-synthetic endoperoxides involves ROS-dependent depolarization of the membrane potential." *Journal of Antimicrobial Chemotherapy*. 69.4. Pg 1005-1016.
- Butler, A.R., Gilbert, B.C., Hulme, P., Irvine, L.R., Renton, L. and Whitwood, A.C. (1998) "EPR evidence for the involvement of free radicals in the iron-catalysed decomposition of qinghaosu (artemisinin) and some derivatives; antimalarial action of some polycyclic endoperoxides." *Free Radical Research*. 28.5. Pg 471-476.
- Muangphrom, P., Seki, H., Fukushima, E.O. and Muranaka, T. (2016) "Artemisinin-based antimalarial research: Application of biotechnology to the production of artemisinin, its mode of action, and the mechanism of resistance of *Plasmodium* parasites." *Journal of Natural Medicines*. 70. Pg 318-334.
- Zhang, Y., Teoh, K., Reed, D., Maes, L., Goossens, A., Olson, D., Ross, A. and Covello, P.S. (2008) "The molecular cloning of artemisinic aldehyde Delta11(13) reductase and its role in glandular trichome-dependent biosynthesis of artemisinin in *Artemisia annua*." *Journal of Biological Chemistry*. 283.31. Pg 21501-21508.
- Kudaksseril, G.J., Lam, L. and Staba, E.J. (1987) "Effect of sterol inhibitors on the incorporation of ¹⁴C-Isopentenyl pyrophosphate into artemisinin by a cell-free system from *Artemisia annua* tissue cultures and plants." *Planta Medica*. 53.3 Pg 280-284.
- Wen, W. and Yu, R. (2011) "Artemisinin biosynthesis and its regulatory enzymes: Progress and perspective." *Pharmacognosy Reviews*. 5.10. Pg 189-194.
- Kim, S-H., Heo, K., Chang, Y-J., Park, S-H. Rhee, S-K. and Kim, S-U. (2006) "Cyclization mechanism of Amorpha-4,11-diene synthase, a key enzyme in artemisinin biosynthesis." *Journal of Natural Products*. 69.5. Pg 758-762.
- Brown, G. D. (2010) "The biosynthesis of artemisinin (qinghaosu) and the phytochemistry of *Artemisia annua* L. (Qinghao)." *Molecules*. 15.11. Pg 7603-7698.
- Qi, X., Chen, X. and Wang, Y. (2014) *Plant metabolomics: methods and applications*. Springer Publishing.
- Czechowski, T., Larson, T.R., Catania, T.M., Harvey, D., Brown, G.D. and Graham, I.A. (2016), "*Artemisia annua* mutant impaired in artemisinin synthesis demonstrates importance of nonenzymatic conversion in terpenoid metabolism." *Proceedings of the National Academy of Sciences U.S.A.* 113.49. Pg 15150-15155.
- Wallaart, T.E., van Uden, W., Lubberink, H.G.M., Woerdenbag, H.J., Pras, N. and Quax, W.J. (1999) "Isolation and identification of dihydroartemisinic acid from *Artemisia annua* and its possible role in the biosynthesis of artemisinin." *Journal of Natural Products*. 62.3. Pg 430-433.
- Wallaart, T.E., Pras, N. and Quax, W.J. (1999) "Isolation and identification of dihydroartemisinic acid hydroperoxide from *Artemisia annua*: A novel biosynthetic precursor of artemisinin." *Journal of Natural Products*. 62.8. Pg 1160-1162.
- Paddon, C.J. and Keasling, J.D. (2014) "Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development." *Nature Reviews Microbiology*. 12. Pg 355-367.
- N.A. (2006) "WHO monograph on good agricultural and collection practices (GACP) for *A. annua* L." *WHO*, WHO. <<http://www.who.int/medicines/publications/traditional/ArtemisiaMonograph.pdf>> (accessed 08/12/2016)
- Simonnet, X., Quennoz, M. and Carlen, C., (2008) "New *Artemisia Annua* hybrids with high artemisinin content." *Acta Horticulturae*. 769. Pg 371-373.
- Delhi, T., Solem, C. and Jensen, P.R. (2012) "Tunable promoters in synthetic and systems biology." *Subcell Biochemistry*. 64. Pg 181-201.
- Jensen, M.K. and Keasling, J.D. (2015) "Recent applications of synthetic biology tools for yeast metabolic engineering." *FEMS Yeast Research*, 15.1. Pg 1-10.
- Jensen, P.R., and Hammer, K. (1998) "The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters." *Applied and Environmental Microbiology*. 61.1. Pg 82-87.
- N.A. (2014) "First antimalarial treatments produced with semisynthetic artemisinin enter market" *Sanofi*. Sanofi. <<http://en.sanofi.com>> (Accessed 02/12/2016)
- Paddon, C.J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D., Polichuk, D., Teoh, K.H., Reed, D.W., Treynor, T., Lenihan, J., Jiang, H., Fleck, M., Bajad, S., Dang, G., Dengrove, D., Diola, D., Dorin, G., Ellens, K.W., Fickes, S., Galazzo, J., Gaucher, S.P., Geistlinger, T., Henry, R., Hepp, M., Horning, T., Iqbal, T., Kizer, L., Lieu, B., Melis, D., Moss, N., Regentin, R., Secrest, S., Tsuruta, H., Vazquez, R., Westblade, L.F., Xu, L., Yu, M., Zhang, Y., Zhao, L., Lievens, J., Covello, P.S., Keasling, J.D., Reiling, K.K., Renninger N.S. and Newman, J.D. (2013) "High-level semi-synthetic production of the potent antimalarial artemisinin." *Nature*. 496.7446. Pg 528-532.
- Westfall, P.J., Pitera, D.J., Lenihan, J.R., Eng, D., Woolard, F.X., Regentin, R., Horning, T., Tsuruta, H., Melis, D.J., Owens, A., Fickes, S., Diola, D., Benjamin, K.R., Keasling, J.D., Leavell, M.D., McPhee, D.J., Renninger, N.S., Newman, J.D. and Paddon, C.J. (2012) "Production of amorphaadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin." *Proceedings of the National Academy of Sciences U.S.A.* 109.3. Pg E111-E118.
- Turconi, J., Griolet, F., Guevel, R., Oddon, G., Villa, R., Geatti, A., Hvala, M., Rossen, K., Goller, R. and Burgard, A. (2014) "Semisynthetic artemisinin, the chemical path to industrial production." *Organic Process Research and Development*. 18.3. Pg 417-422.
- Amara, Z., Bellamy, J., Horvath, R., Miller, S., Beeby, A., Burgard, A., Rossen, K., Poliakov, M. and George, M. (2015) "Applying green chemistry to the photochemical route to artemisinin." *Nature Chemistry*. 7.6. Pg 489-495.
- Singh, D., McPhee, D., Paddon, C.J., Cherry, J., Maurya, G., Mahale, G., Patel, Y., Kumar, N., Singh, S., Sharma, B., Kushwaha, L., Singh, S. and Kumar, A. (2017) "Amalgamation of synthetic biology and chemistry for high-throughput nonconventional synthesis of the antimalarial drug artemisinin." *Organic Process Research and Development*. 21.4. Pg 551-558.
- Feth, M.P., Rossen, K. and Burgard, A. (2013) "Pilot plant PAT approach for the diastereoselective diimide reduction of artemisinic acid." *Organic Process Research and Development*. 17.2. Pg 282-293.
- Seeberger, P.H. and Lévesque, F. (2012) "Continuous-flow synthesis of the anti-malaria drug artemisinin." *Angewandte Chemie International Edition*. 51.7. Pg 1706-1709.
- Seeberger, P.H., Lévesque, F. and Kopetzki, D. (2013) "A continuous-flow process for the synthesis of artemisinin." *Chemistry – A European Journal*. 19.17. Pg 5450-5456.

Lactate Promotes Survival of Peripheral Blood Mononuclear Cells by Decreasing Apoptotic Events

BY SARAH L.C. ELSASER^{1,2,3,4}, FIONA F. VICKERS^{1,2},

CATALINA M. CARVAJAL GONCZI^{1,2,4}, PETER J. DARLINGTON^{1,2,4}

¹CONCORDIA UNIVERSITY, MONTREAL, QUEBEC, CANADA.

²DEPARTMENT OF EXERCISE SCIENCE AND DEPARTMENT OF BIOLOGY

³SCIENCE COLLEGE

⁴PERFORM CENTRE AND CENTRE FOR STRUCTURAL AND FUNCTIONAL GENOMICS

INTRODUCTION

Moderate exercise can increase the efficiency of the immune system and reduce the risk of infection.¹ In contrast, strenuous exercise, such as intense and prolonged training, can inhibit immune function and increase the risk for upper respiratory tract infections.² Strenuous exercise can cause lymphopenia, a reduction in the proportions of white blood cells, depending on the type and intensity of exercise.³ Lymphopenia could be due to decreased flow of lymph into blood, or by immune cell death by necrosis or apoptosis.⁴ The biochemical mechanism of action by which exercise modifies the immune system is not completely understood. During anaerobic exercise, working muscles utilize glycogen, leading to the formation of lactate, which can spill over into the blood. Traditionally, lactate has been viewed as a by-product of exercise that can lead to fatigue. In 2007, however, lactate was found to actually provide protective effects to rat skeletal muscles, suggesting that it can enhance cell survival.⁵

Cells typically die by two ways: necrosis or apoptosis. Necrosis is characterized as a passive cell death following environmental perturbations, whereas apoptosis is a regulated process of 'programmed' cell death. Lymphocyte necrosis and apoptosis were higher immediately following a 24-hour period after a high intensity treadmill exercise test, as they were compared to the values obtained after an exercise of moderate intensity.^{6,7}

We reasoned that lactate could be one of the factors acting as a bridge between exercise and the immune system. Lactic acid produced during exercise is converted in the form of its conjugated base lactate. Lactate accumulates in muscles where it is mostly consumed, and some enters the circulation. In an oxygen-poor environment, such as

in muscle cells during exercise, lactate can be produced from pyruvate catalysed by the enzyme lactate dehydrogenase (LDH) where pyruvate + NADH + H⁺ → lactate + NAD⁺. If lactate builds up in muscles, it will stimulate pain sensation and promote damage to the tissue, and it will also cause lactate levels to increase in the blood. The clearance of blood lactate can be improved by endurance training.⁸

Lactate is metabolized by the heart, neurons in the brain, and hepatic cells, which can convert lactate into glucose via the Cori cycle.⁹ However, what is the effect of blood lactate on the immune system? In the current study, lactate was explored as a possible chemical factor linking together effects of exercise and immune responses. Peripheral blood mononuclear cells (PBMCs) were used for this study as a source of human immune cells. This blood fraction, containing a mix of lymphocytes and monocytes that belong to the white blood cell family, is commonly used for in vitro research on the immune system.

MATERIALS AND METHODS

PBMC ISOLATION AND CELL CULTURE

The study was conducted using human PBMCs. Whole human blood was obtained by venipuncture (approved by the Concordia University Human Research Ethics Committee). The Blood samples were diluted in a solution of 1X phosphate buffered saline (PBS) and a ratio of 1:2 Ficoll to separate blood into different layers. After 30 minutes of centrifugation at 1800 rpm (700 X g) on the Thermo Scientific CL10 series with a 0-G26/1 swing-out rotor, four layers were observed: PBS with plasma as the top layer, followed by a PBMCs layer, then Ficoll, and finally erythrocytes as bottom layer. The

PBMCs layer was carefully removed with a transfer pipette, placed into fresh conical tube, diluted with PBS, and centrifuged for 15 minutes at 1500 rpm (480 X g) with the same centrifuge. PBMCs formed a pellet, and can be used after washing twice with PBS at 1200 rpm (310 X g).¹⁰

Six different media were prepared from a stock of Roswell Park Memorial Institute media (RPMI; Wisent Bioproducts) with L-glutamine, penicillin-streptomycin and 2mg/mL glucose. The first medium was a lactate-free medium. The other five media contained lactate at 0.5mM, 1.0mM, 3.5mM, 7mM, and 14mM, respectively. Lactate was purchased from SIGMA Aldrich. The pH values measured for the media containing lactate at 0mM, 0.5mM, 1.0mM, 3.5mM, 7.0mM, and 14mM were, respectively, 7.56, 7.51, 7.49, 7.42, 7.35, and 7.08. The PBMCs (at 4.0x10⁶ cells/mL) were added to each medium in a 96 well round-bottom sterile plate. Two duplicates were made to account for within-test variance.

FLOW CYTOMETRY AND APOPTOSIS MEASUREMENTS

The culture plate containing PBMCs was incubated at 37°C and cell counting was performed by flow cytometry using fluorescent dyes, including a propidium iodide analogue (7AAD) and annexin V. Percentages for live, necrotic, and apoptotic cells were obtained using the Acurri BD C6 Plus software by placing a quadrant that falls between the populations of cells. Samples without any dye added were used as a negative control to set the quadrants correctly.

STATISTICS

The data was processed for removal of within subject variations, according to Loftus et al.¹¹ ANOVA analysis was done on each data set from day 1, day 2, and day 3, and a post-hoc test (Tukey's multiple comparison) test was performed.

RESULTS

Samples of PBMCs were exposed in vitro to a range of lactate concentrations intended to simulate biologically relevant blood lactate concentrations. The samples were analysed each day for a total of three days to determine the percentages for live, necrotic, and apoptotic cells using flow cytometry (Figure 1). 7AAD is a measure that indicates necrosis and late apoptosis, and annexin V is a measure of early apoptosis. If a cell has holes in its membrane, then it will be positive for 7AAD and positive for annexin V. Thus, necrosis is 7AAD-positive and annexin V positive, and apoptosis is 7AAD-negative annexin V positive, as shown in the figure 1 quadrant analysis.

On day 1 of the analysis, which corresponded to 24 hours after starting the incubation, there was no significant difference in the percentage of live cells in the cultures (Figure 2A). On day 2, the survival was still the same for all lactate concentrations, although the overall amount of live cells appeared to be lower than on day 1, suggesting that the cells were slowly dying in culture (Figure 2B). By day 3, there were significantly higher survival rates in the 3.5mM and 7.0mM lactate concentrations as compared to the culture without lactate (Figure 2C). Necrosis was not significantly different between the lactate concentrations at any of the times, although there was an overall trend of increasing necrosis from day 1 to day 2, and from day 2 to day 3 (Figure 2D-F). Interestingly on day 2, the percentage of apoptosis was lower for the 3.5mM lactate as compared to no lactate, and on day 3, apoptosis was lower for the 3.5mM and 7.0mM lactate compared to the control group without lactate (Figure 2G-I).

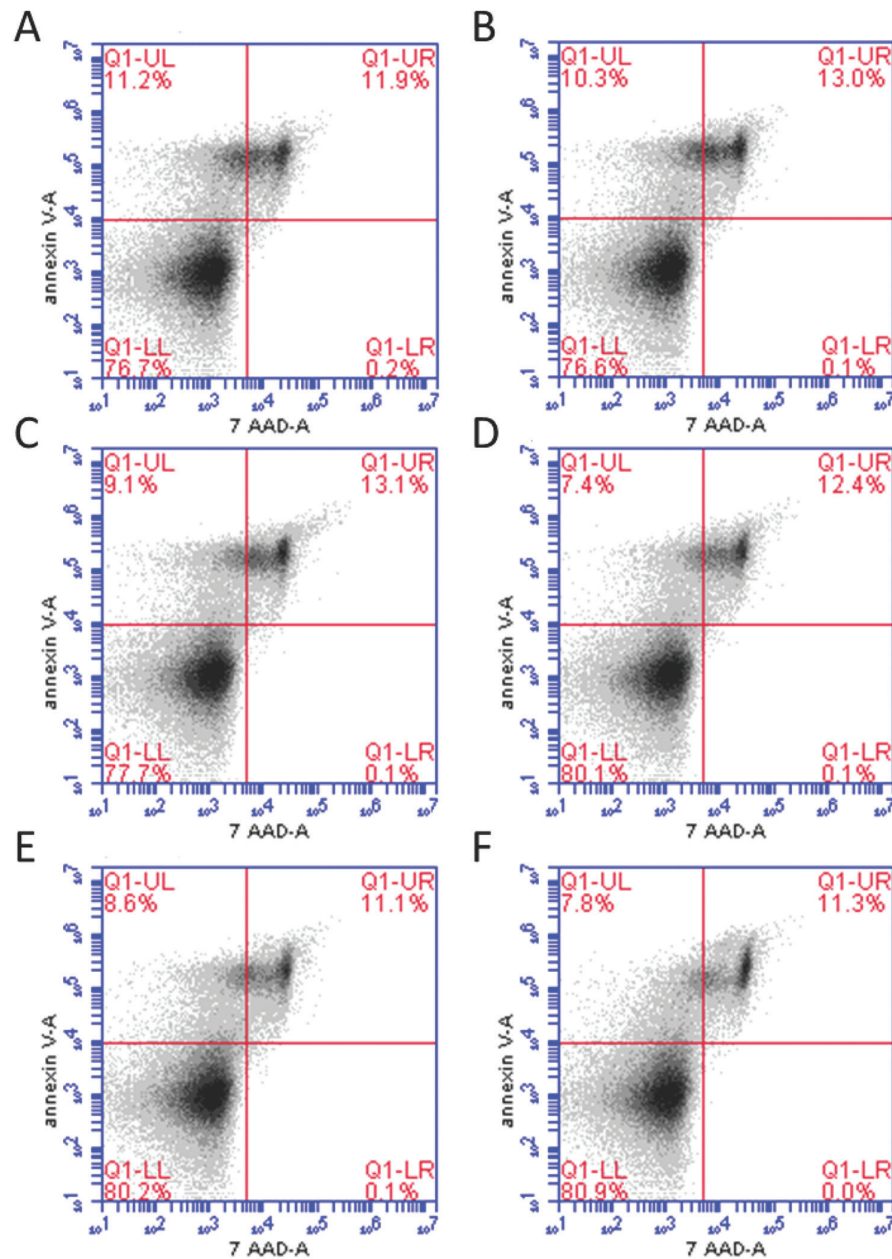
Looking at trends for peak values on each day, the highest percentage of live cells was present in the 7mM lactate medium for day 1, the 3.5mM lactate medium for day 2, and the 7mM lactate medium for day 3 (Figure 2A-C). The highest percentage of necrotic cells was present for the 1mM lactate treatment for day 1, the 1mM lactate for day 2, and the 14mM lactate for day 3 (Figure 2D-F). The highest percentage of apoptotic cells was present in the 0.5mM lactate medium for day 1 and the 0mM for day 2 and day 3 (Figure 2G-I). This analysis is consistent with the notion that cells are surviving better at 3.5mM and 7mM lactate compared to the lower values of lactate or the higher value of 14mM. The enhanced survival is due to less apoptosis occurring when PBMCs have 3.5mM or 7.0mM of lactate. With 14mM of lactate, the PBMCs had a non-significant trend towards surviving more and having less apoptosis than the 0mM condition.

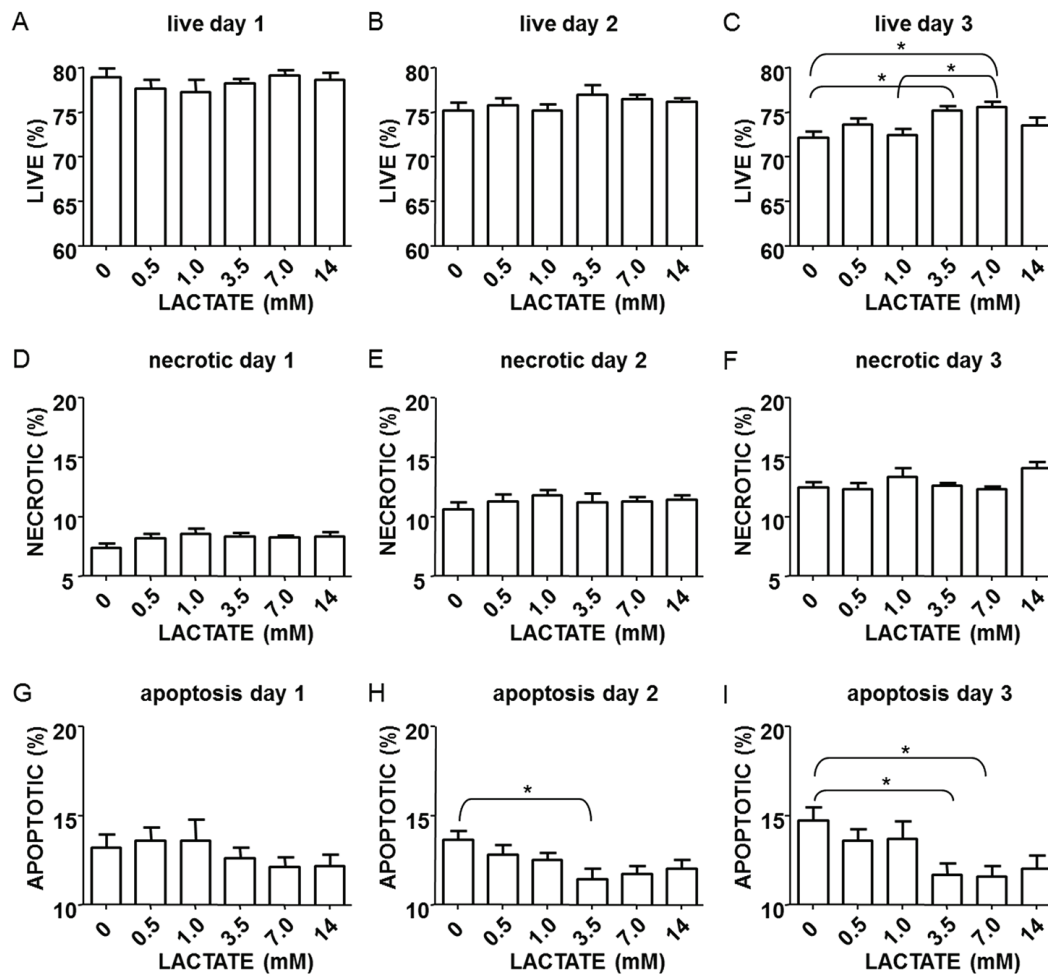
Figure 1: Example flow cytometry data and quadrant analysis

PBMCs were untreated (A), or treated with lactate at 0.5mM (B), 1.0 mM (C), 3.5 mM (D), 7.0 mM (E), and 14 mM (F). Shown is example data of a day 3 analysis. Cells were stained with annexin V and 7AAD, then analysed by flow cytometry. Each plot is divided into four quadrants by the cross, and the percentage of cells that is in each quadrant is shown. The lower left (LL) quadrants are live cells, the upper left (UL) are apoptotic cells, and the upper right (UR) are necrotic cells. Data is representative of 5 experiments.

FIGURE 2: Summary of results from PBMCs treated with various concentrations of lactate

The percentage of cells is shown for live, necrotic, and apoptotic cells, as determined based on the quadrant analysis criteria in Figure 1. The data is pooled from 5 healthy people's PBMC samples. The samples had been treated in vitro with lactate for 1 day, 2 days, or 3 days. *p<0.05





DISCUSSION

The underlying mechanism of action by which exercise affects the immune system remains unclear. In this study, blood lactate was explored as a factor that may affect lymphocyte survival using an in vitro modelling approach. The amount of lactate we chose to add to the PBMCs was based on biologically relevant values of lactate concentration in the blood that has been recorded during various levels and modes of exercise. The values represent increasing intensities of exercise. The 0.5mM and 1.0mM values correspond to plasma lactate concentrations when a person is at rest.¹² The small amount of blood lactate at rest is produced by red blood cells from anaerobic respiration.¹³ The 3.5mM lactate value reflects lactate concentrations measured at a work rate corresponding to low-intensity exercise.¹⁴ The 7.0mM lactate value corresponds to the amount of lactate present in the blood 5 minutes after exercising at 60% VO_2 max, equivalent to a medium intensity.¹⁵ The 14.0mM lactate corresponds to the amount present in the blood after exercising at 90% VO_2

max, equivalent to a high-intensity exercise involving arm cranking or knee extension in healthy men.^{14,16} The 14.0mM concentration also represents a high value recorded after maximum-intensity treadmill running by trained men subjects aged 22 to 29 years.¹⁷ By recreating these values of lactate, we modelled in vitro what effect they have on lymphocytes. This reductionist approach is advantageous in that we can control the amount of lactate precisely and minimize other confounding variables that exist when a person is exercising.

Our results demonstrate that immune cells survive better in vitro at the higher lactate concentrations that represent low to medium intensity exercise (3.5mM, 7.0mM), and high intensity exercise to an extent (14.0mM). The lower lactate concentrations that represent resting levels of lactate had little effect on PBMC survival. The enhanced survival with the higher lactate concentrations was due to decreased apoptosis, rather than necrosis. It was surprising that PBMCs tolerated the 14mM lactate condition well. Although

there was a trend of more necrosis on day 3 with the 14mM lactate, the live cells were similar to the control with no lactate.

Being an acid in fluid means that lactate will lower the pH of the growth media. Despite the high concentration (14mM) of lactate added to the media, we measured a neutral pH (7.0), as compared to the control media, which was measured at a pH of 7.5. The minimal difference in pH in the various lactate medias we prepared is due to the strong bicarbonate buffer that is present in the media purchased from the manufacturer. Blood also contains a bicarbonate buffer, which can counteract acids, so the media we used closely matches that of blood.

The fact that lactate promoted cell survival suggests that it may provide a source of carbohydrate which the cells can use for energy. In theory, immune cells would require a lactate transporter in their membranes and lactate dehydrogenase to convert the lactate back into glucose. Monocarboxylate transporters are found in human lymphocytes¹⁸. More research is required to determine the extent to which white blood cells might use lactate for energy.

There were some limitations to this study. The PBMCs used were a mixture of cell types, so it is not possible to say definitively that the cell measures are coming from a particular type of mononuclear cell or all mononuclear cell types, including T cell, B cell, monocytes, NK cells, and other rare subsets. Cell sorting would be one way to address this issue in future experiments. Another limitation that is inherent to reductionist experiments is that it does not necessarily translate into an actual exercise paradigm. While these experiments were done in vitro, they do shed light on the in vivo interaction between the immune system and exercise. We exposed PBMCs to lactate and examined the effects over a period of three days. We chose this time frame because apoptosis typically occurs over the course of approximately three days, and we wanted to ensure we captured any effect on cell survival.

In conclusion, lactate at levels reflecting low to moderate or strenuous exercise promoted survival of PBMCs. Considering the evidence that strenuous exercise induces apoptosis, it seems unlikely that lactate is the reason for apoptosis in the context of exercise. On the contrary, immune cells appeared to benefit from lactate. In fact, it could be that lactate production is one of the reasons that moderate exercise enhances the immune system.

ACKNOWLEDGEMENTS

This work was funded by NSERC Discovery grant (RGPIN 418522-2013). Thank you to Mahdieh Tabatabaei Shafiei for proofreading the manuscript.

REFERENCES

- ¹Terra, R., Silva, S., Pinto, V. and Dutra, P. (2012). "Effect of exercise on immune system: response, adaptation and cell signalling." *Rev Bras Med Esporte*. Vol 18. No 3. Pg 208-214.
- ²Walsh, N., Gleeson, M., Shephard, R., Woods, J., Bishop, N., Fleshner, M., Green, C., Pedersen, B., Hoffman-Goetz, L., Rogers, C., Northoff, H., Abbasi, A. and Simon, P. (2011). "Position statement. Part one: Immune function and exercise." *Exercise Immunology Review*. 17.6. Pg 63.
- ³Phaneuf, S. and Leeuwenburgh, C. (2001). "Apoptosis and exercise." *Medicine and Science in Sports and Exercise*. 33.3. Pg 393-396.
- ⁴Fink, S. and Cookson, B. (2005). "Apoptosis, pyroptosis, and necrosis: mechanistic description of dead and dying eukaryotic cells." *Infection and immunity*. 73. 4.
- ⁵Paoli, F., Overgaard, K., Pedersen, T. and Nielsen, O. (2007). "Additive protective effects of the addition of lactic acid and adrenaline on excitability and force in isolated rat skeletal muscle depressed by elevated extracellular K⁺." *The Journal of Physiology*. 581. Pg 829-839.
- ⁶Mooren, F., Blöming, D., Lechtermann, A., Lerch and M., Völker, K. (2002). "Lymphocyte apoptosis after exhaustive and moderate exercise." *Journal of Applied Physiology*. 93. 1. Pg 147-153.
- ⁷Mooren, F. and Krüger, K. (2015). "Apoptotic lymphocytes induce progenitor cell mobilization after exercise." *Journal of Applied Physiology*. 119. 2. Pg 135-139.

⁸Fukuba, Y., Walsh, M., Morton, R., Cameron, B., Kenny, C. and Banister, E. (1999). "Effect of endurance training on blood lactate clearance after maximal exercise." *Journal of Sports Science*. 17.3. Pg 239-48.

⁹Wyss, M., Jolivet, R., Buck, A., Magistretti, P. and Weber, B. (2011). "In Vivo Evidence for Lactate as Neuronal energy source." *The Journal of Neuroscience*. 31.20. Pg 7477-7485.

¹⁰Tabatabaei Shafiei, M., Carvajal Gonczi, C., Samiur Rahman, M., East, A., François, J. and Darlington, P. (2014). "Detecting glycogen in peripheral blood mononuclear cells with periodic acid schiff staining." *Journal of Visualized Experiments*. 94. doi:10.3791/52199.

¹¹Loftus, G. and Masson, M. (1994). "Using confidence intervals in within-subjects designs." *Psychonomic Bulletin & Review*. 1.4. Pg 476-490.

¹²Phypers, B. and Pierce, J. (2006). "Lactate physiology in health and disease." *Continuing Education in Anesthesia, Critical Care & Pain*. 6.3. Pg 128-132.

¹³Hildebrand, A., Lormes, W., Emmert, J. and Steinacker, J. (2000). "Lactate concentration in plasma and red blood cells during incremental exercise." *International Journal of Sports Medicine*. 21.7. Pg 463-468.

¹⁴Goodwin, M., Harris, J., Hernandez, A. and Gladden, L. (2007). "Blood Lactate Measurements and Analysis during Exercise: A Guide to Clinicians." *Journal of Diabetes Science and Technology*. 1.4. Pg 558-569.

¹⁵Bangsbo, J., Aagaard, T., Olsen, M., Kiens, B., Turcotte, L. and Richter, E. (1995). "Lactate and H⁺ uptake in inactive muscles during intense exercise in man." *Journal of Physiology*. 488.1. Pg 219-229.

¹⁶Gass, G., Rogers, S. and Mitchell, R. (1981). "Blood Lactate Concentration Following Maximum Exercise In Trained Subjects". *British Journal of Sports Medicine*. 15.3. Pg 172-176.

¹⁷Hertz, L. and Diemel, G. (2005). "Lactate transport and transporters: general principles and functional roles in brain cells." *Journal of Neuroscience Research*. 79.1-2. Pg 11-8.

¹⁸Daberkow RL., White BR., Cederber RA., Griffin JB. and Zemleni J. (2003). "Monocarboxylate transporter 1 mediates biotin uptake in human peripheral blood mononuclear cells." *The Journal of Nutrition*. Vol. 133, no.9, 2703-2706.

Whole Exome Sequencing and Analysis to Rule Out Three Bone-Expressed Genes as Candidates for Split Hand Foot Malformation

C. RONNIE FUNK,¹ MELANIE M. MAY,² ANNA V. BLENDA, PH.D.,³ & CHARLES E. SCHWARTZ, PH.D.²

¹ ERSKINE COLLEGE, ² GREENWOOD GENETIC CENTER, & ³ USC SCHOOL OF MEDICINE GREENVILLE

Abstract

Family studies are useful in the genetic analysis of rare conditions. In this study, a rare autosomal dominant disease, called split hand foot malformation (SHFM), was analyzed in a family. Mutations in the known SHFM genes are thought to abrogate development of the central rays of the hands and feet through various mechanisms. A typical SHFM phenotype manifests as syndactyly, median clefts of the hands and feet, and/or aplasia or hypoplasia of the phalanges, metacarpals, and metatarsals. Although family members presented with the SHFM phenotype, mutations were absent in the nine loci associated with nonsyndromal SHFM, while patient history and presentation eliminated syndromal SHFM as a potential diagnosis. In order to determine a hypothesized genetic cause of SHFM in this family, genomic DNA of four family members was subjected to whole exome sequencing (WES) to screen all protein-encoding regions of the DNA. The sequencing data were filtered to identify gene alterations that were novel, and not present in databases of sequence variants. Based on the analysis, mutations in three genes, each expressed in bone, *MAP1b*, *NEK1*, and *CHD6* were selected for further investigation based on known biological functions. While this work was underway, twins – one of whom was clearly affected – were born to an affected father. The extension of the pedigree enabled segregation analysis based on DNA sequencing to observe whether mutations in *MAP1b*, *NEK1*, and *CHD6* were present in the affected newborn twin. Through this methodology, none of the three gene alterations were associated with the SHFM phenotype in the family. This exposition conveys a methodology for analyzing a rare genetic disorder with the dual goals of strengthening future study designs and communicating the value of confirming WES data in order to prevent incorrectly counseling patients and their families on rare genetic conditions.

INTRODUCTION

Patients with Split Hand Foot Malformation (SHFM) may present with polydactyly, syndactyly, median clefts, hypoplasia, or hyperplasia of the digits of the hands and feet.¹ Abrogation of central-ray formation in developing limbs is thought to cause SHFM.² Multiple genetic mutations can cause SHFM through putative alteration of gene expression in the proliferating zone of the apical ectodermal ridge in a developing limb bud.¹

In some patients, SHFM is present as an isolated condition without other abnormalities, and these patients are properly diagnosed with nonsyndromal SHFM.³ In other patients, SHFM occurs along with additional abnormalities such as long-bone deficiency or cleft lip/palate, and these cases are known as syndromal SHFM.^{4,5} In both nonsyndromal and syndromal SHFM, penetrance and expressivity of the malformation varies.^{3,6} Thus, penetrance and expressivity need to be considered to accurately interpret segregation analyses.

Family K6821 presented a nonsyndromal SHFM phenotype. The family elected to undergo genetic testing to determine the cause of their condition. Through a variety of screening and diagnostic procedures, this family was found to lack mutations in any of the six unique chromosomal loci associated

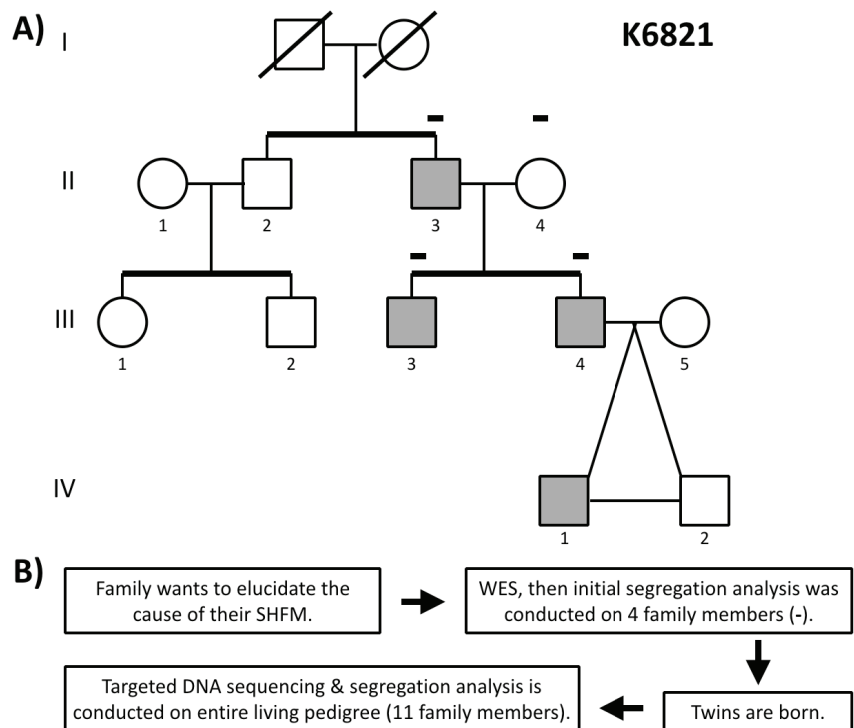


Figure 1: A) Pedigree of family K6821. Individuals with the SHFM phenotype are shaded gray. Square = male; circle = female, triangle = identical twins, and a bar over square or circle denotes Whole Exome Sequencing (WES).

B) Chronology of events in this study.

Table 1: RNA primer sequences used to amplify MAP1b Exon 4, NEK1 Exon 31, and CHD6 Exon 23.

Gene	¹ Forward Primer (5' to 3')	² Reverse Primer (5' to 3')
M13 Tag	TTGTAAAACGACGGCCAGTG	CACACAGGAAACAGCTATGACCATG
MAP1b Exon 4	M13F-CTAGCCCTGTCTGAAGGTG	M13R-TTCTTTTGTCTCCCCCAGTG
NEK1 Exon 31	M13F-TTGTGTTACATACTGTGCCTTATCAA	M13R-TGCAAAGGATTCTTCTGGTG
CHD6 Exon 23	M13F-CCACTTGCCTATGCTGTCT	M13R-CAGAAGTAGACAAGAGCATGAA

¹Forward primers target intronic DNA upstream of target genes.

²Reverse primers target intronic DNA downstream of target genes.

with nonsyndromal SHFM (7q21.2-q21.3, Xq26, 10q24, 3q27, 2q31, and 12q31).¹ Furthermore, the family lacked a rearrangement at the three chromosomal loci associated with SHFM or SHFM with long-bone deficiency (1q42.2-q43, 6q14.1, and 17p13.3).¹ In summation, these data indicated the family lacked a chromosomal rearrangement association of known nonsyndromal SHFM genes.

An investigation was designed and initiated with the goal of elucidating the unknown molecular etiology of SHFM in this family. Due to the family history of SHFM, the underlying etiology was hypothesized to involve genetics, rather than being solely environmental. A genetic mutation in known exonic DNA was hypothesized to underlie SHFM in this family. Thus, whole exome sequencing (WES) was selected as the genetic test of choice. WES enables analysis of all the variation present in known exonic regions of DNA. From the wealth of variants detected by WES, bioinformatic analysis of the WES data was conducted with the goal of identifying novel candidate SHFM genes for further investigation based on the putative biologic functions of the encoded proteins. These candidate genes were then studied further using DNA sequencing and segregation analyses. In this exposition, the methods and thought processes used to systematically search for a novel cause of a rare genetic condition are discussed, with emphasis upon the value of an extended pedigree when studying a genetic condition.

MATERIALS AND METHODS

Whole Exome Sequencing

Initial WES was conducted by Beijing Genomics Institute on four affected individuals (Figure 1A). Genomic DNA from all members of family K6821 was isolated from blood samples after informed consent was obtained. The chronology of events is shown in Figure 1B to highlight that WES and an initial segregation analysis were conducted before the birth of a set of identical twins.

Bioinformatic Analysis

The WES data from four individuals in family K6821, consisting of approximately 100,000 single nucleotide variants, were im-

ported into Ingenuity Variant Analysis (IVA). IVA was used to analyze the variants and calculate allele frequencies using the 1000 Genomes database. IVA was also utilized to predict whether the variants were likely to be pathogenic. The screened data were imported to Excel and selected for variants that were heterozygous in the three affected individuals (II-3, III-3, III-4) but not present in the single unaffected individual (II-4). Rare variants (allele frequency <1%) were then screened for nonsynonymous variants. Using these data, the Sorting Intolerant from Tolerant (SIFT) program, which sorts intolerant from tolerant amino acid substitutions; and Polyphen, a program that predicts effects of amino acid substitution, were used to predict pathogenic variants for further analyses. The data were then filtered to only include alterations identified in genes with expression in bone. From this analysis, heterozygous mutation in three bone-expressed genes was identified: *MAP1b*, *NEK1*, and *CHD6*.

Gene-Specific Sequencing of MAP1b, NEK1, & CHD6

Amplifications of *MAP1b* exon 4, *NEK1* exon 31, and *CHD6* exon 23 were conducted using intronic RNA primers conjugated to M13 tags (Table 1).

To amplify *MAP1b* exon 4, *NEK1* exon 31, and *CHD6* exon 23, annealing temperatures were optimized by gradient PCR. Using primers shown in Table 1, each exon was amplified by initially denaturing the sample for 5 minutes at 95°C, then conducting 35 cycles of denaturation (95°C, 30s), annealing (optimized temperature, 30s), and adding extension (72°C, 40s), with a final extension of 72°C for 5 minutes. GoTaq® DNA Polymerase and 5X Green GoTaq® Reaction Buffer (Promega, Madison, WI) were used in all DNA amplifications. Electrophoresis using 2% agarose gels was then conducted to qualitatively assess DNA amplification.

ExoSAP-IT (Affymetrix, Santa Clara, CA) enzyme cocktail was added to the amplified DNA product, and the solution was heated to 37°C to degrade the primers and dNTPs. Then the reaction mixture was heated to 80°C to denature the enzyme without reaching the T_m of the duplex DNA products. This purified sample, devoid of

primers, dNTPs, and other short DNA sequences, was labeled with BigDye® (Thermo Fisher Scientific, Grand Island, NY) by first disrupting the hydrogen bonds in the duplex DNA at 96°C for 1 minute, then subjecting the samples to 25 cycles consisting of denaturation at 96°C for 10 seconds, with 5 seconds of labelling at 50°C and 1 minute 15 seconds of extension at 60°C. BigDye® Terminator v1.1 and v3.1 5X Sequencing Buffer (Thermo Fisher Scientific, Grand Island, NY) were used to stabilize the sequencing reagents. The DNA product labelled with BigDye® was purified using Qiagen DYE-EX 96-well processing plates (Qiagen, Venlo, Limburg). Capillary electrophoresis using a 3730 DNA Analyzer (ThermoFischer Scientific, Grand Island, NY) was performed.

RESULTS

WES, followed by bioinformatic analysis, identified single nucleotide polymorphisms (SNPs) in *MAP1b*, *NEK1*, and *CHD6* as candidate mutations causing the SHFM phenotype in family K6821. Sequencing of exon 4 of *MAP1b* identified an adenosine to guanine change, c.437 A>G, in three affected individuals: II-3, III-3, III-4 (Figure 2). This DNA variation was predicted to encode an amino acid change from asparagine to serine at residue 146 (p.A146S). Each family member with SHFM was heterozygous for the c.437G>A change except for the affected newborn twin IV-1 (Figure 2). These data indicated that the missense mutation in *MAP1b* does not co-segregate with SHFM in K6821.

DNA sequencing data for *NEK1* (Figure 3) revealed that individuals II-3, III-3, and III-4 were heterozygous for the mutation at c.2975-2 A>T. Furthermore, the polymorphism was not present in individuals lacking the SHFM phenotype. Analysis of the newborn twins revealed lack of segregation of this alteration in the affected twin, IV-1, thereby demonstrating that this mutation in *NEK1* is not associated with SHFM in this family.

A third mutation identified by the bioinformatic analysis of the WES data was a

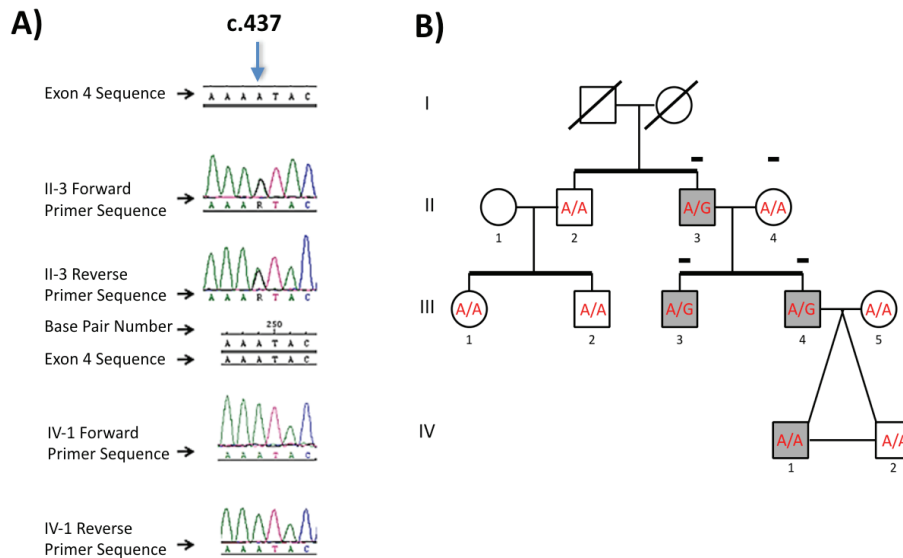


Figure 2: A) DNA sequence data for II-3 (top) and IV-1 (bottom) showing change c. 437 A>G (blue arrow). R denotes A or G. B) DNA sequence data generated for members of K6821 at c. 437 (shown in red). Individuals exhibiting the SHFM phenotype are shaded gray.

missense mutation in *CHD6*. Sequencing showed a change from cytosine to guanine, c.3675C>G. This was predicted to change the amino acid sequence at residue 1225 from histidine to glutamine (p.H1225Q) within exon 23 of *CHD6* isoform 1. Segregation analysis, shown in Figure 4, presented a potential case of variable penetrance in the family until DNA from the affected newborn twin was sequenced and found to lack the c.3675 C>G mutation. Additionally, his unaffected mother who married into the family had the alteration, thereby effectively eliminating the change in *CHD6* gene as pathogenic.

DISCUSSION

Once it became apparent that members of family K6821 lacked mutations in the known SHFM genes, whole exome sequencing (WES) was conducted on four individuals in the family. WES is a useful tool in the study of rare diseases following a Mendelian inheritance pattern because it is the most efficient means for detecting point mutations in all of an individual's genes. Once WES was complete, bioinformatic analysis was performed to analyze variants of unknown significance (VUS) detected by WES. The bioinformatic analysis systematically narrowed down the variants to form a list that could potentially explain the SHFM in family K6821. As part of this analysis, genes with expression in bone were selected, with *MAP1b*, *NEK1*, and *CHD6* hypothesized to be the most significant due to reported

biologic functions of the encoded proteins. The normal chromosomal loci for *MAP1b*, *NEK1* and *CHD6*, are shown in Figure 5.

The reported biological functions of the

three genes are summarized in Table 2.

A limitation of this study is that the SHFM causative gene in this family could be expressed in a different tissue other than bone. In fact, many of the mutations known to cause SHFM affect proteins expressed in the apical ectodermal ridge of a developing limb bud.¹

After systematically searching the literature, it remained unclear as to whether *MAP1b*, *NEK1*, and *CHD6* are expressed in the AER during limb development. Furthermore, even if the causative gene in this family were expressed in bone, another limitation of this study is that genes expressed below detectable levels could have been excluded by the bioinformatic analysis. Despite the limitations of this analysis, the WES data was useful for distinguishing variants in the family's DNA for further study.

Confirmation of the variants identified by WES was also necessary, and Sanger sequencing was utilized to this accomplish this end. A possible limitation of WES is the potential for limited coverage of the sequence containing the causal variant.¹² Sanger sequencing confirmed the results of WES, and initial segregation analyses provided evidence suggesting the three variants could be associated with the SHFM phenotype. Once the twins (IV-1, IV-2, Figure 1) were born, the extension of the pedigree offered a valuable chance to see whether the mutations in *MAP1b*, *NEK1*, and *CHD6* contin-

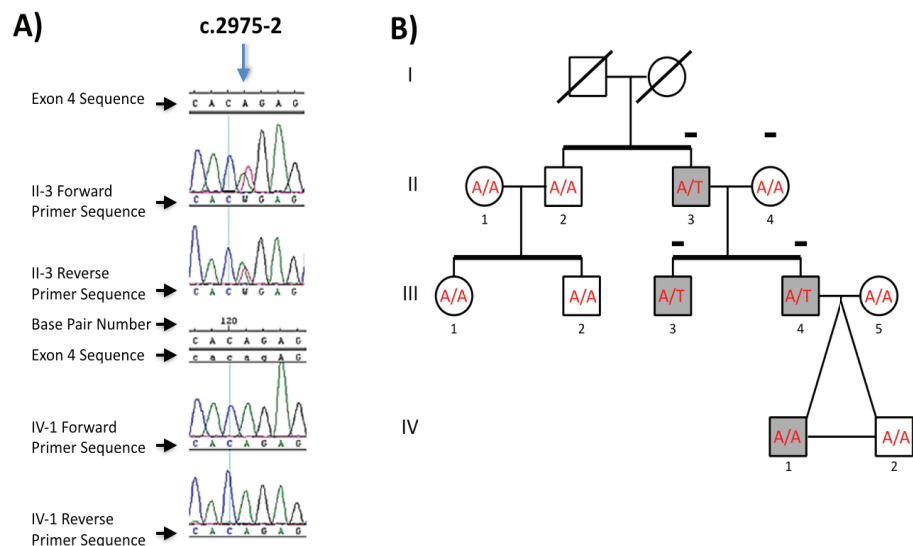


Figure 3: A) DNA sequence data for II-3 (top) and IV-1 (bottom) showing change c. 2975-2 A>T, w (blue arrow). W denotes T or A. B) DNA sequence data generated for members of K6821 at c. 2975-2 A>T (shown in red). Individuals exhibiting the SHFM phenotype are shaded gray.

Table 2: Reported functions of MAP1b, NEK1, and CHD6 proteins that were hypothesized to be significant to the etiology of SHFM.

Gene Product	Overview of Known Function
MAP1b	Regulation of cytoskeletal dynamics and outgrowth. ⁸
NEK1	Regulation of mitotic and meiotic signal transduction cascades. ^{9,10}
CHD6	Alteration of chromatin structure to establish cell type-specific patterns of gene expression. ¹¹

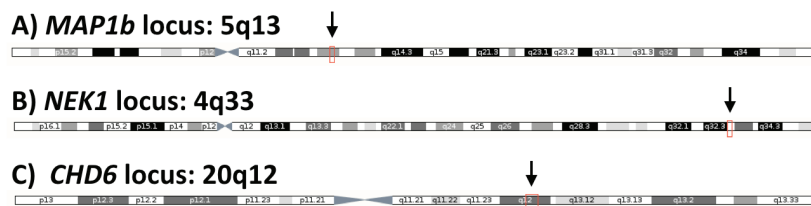


Figure 5: Red boxes, emphasized by arrows, denote the chromosomal loci of *MAP1b*, *NEK1*, and *CHD6* genes. Images obtained from Ensembl database⁷

used to segregate with the SHFM phenotype. Discussion of each of these individual genes follows.

Before the birth of the twins, segregation analysis showed that the mutation in the *MAP1b* gene was present only in affected family members, which suggested this variant could underlie the SHFM phenotype. *MAP1b* is located on the forward strand of chromosome 5 at 5q13 (Figure 5). Microtubule-associated protein 1b (MAP1b) has a function which we hypothesized could be relevant to the SHFM phenotype. MAP1b regulates microtubule dynamics, which is important for axonal outgrowth and cytoskel-

etal growth.^{13,14}

Additionally, MAP1b binds microtubules in a phosphorylation-state-dependent manner when phosphorylated at a serine/threonine kinase specific site.^{8,15,16} We hypothesized that a mutation in a phosphorylated residue could alter MAP1b protein function. The missense mutation present in the MAP1b protein detected was p.N146S. This acquisition of a phosphorylatable serine residue could have enhanced the ability of MAP1b protein to phosphorylate substrates and/or bind microtubules leading to abrogation of central ray formation.

The presence of the missense mutation

in the MAP1b protein in affected individuals would strengthen the gene's candidacy as a gene underlying SHFM. However, the data presented in Figure 2 show that the *MAP1b* mutation is not present in the affected newborn twin, IV-1. Therefore, the extended segregation analysis decreased the likelihood that the *MAP1b* missense mutation is a risk factor for SHFM in family K6821.

In a similar scenario, prior to the birth of the twins, segregation analysis of the mutation, c.2975-2A>T, in the *NEK1* gene showed that mutation was only present in affected family members. NEK1 protein also has a function we hypothesized could be relevant to the SHFM phenotype. Never in Mitosis Gene A related kinase-1 (NEK1) is located on the reverse strand of chromosome 4 at 4q33 (Figure 5) and is a kinase with dual-specificity, since it can phosphorylate either serine/threonine or tyrosine residues on target substrates after appropriate autophosphorylation.^{17,18} This dual-specificity provides unique functions such as regulation of mitotic and meiotic signal transduction cascades.^{9,10}

Recently, the NEK1 protein was demonstrated to mediate S-phase progression by assisting in the loading of replication factors onto chromatin.¹⁹ In K6821, we hypothesized a guanine to cytosine variation two bases before a 5' splice site of intron 30 (c.2975-2 A>T) caused an alteration in a splice site between intron 30 and exon 31, ultimately altering protein structure and function. Although the mutation in *NEK1* was present in patients II-3, III-3, and III-4, it was not found in the affected newborn twin, IV-1 (Figure 3B). These data suggest that the variation seen in *NEK1* is not responsible for the SHFM phenotype in family K6821.

The initial segregation analysis conducted for the c.367C>G mutation in the *CHD6* gene presented the possibility of this variant causing SHFM, but with incomplete penetrance. The *CHD6* gene is located at 20q12 on the reverse strand (Figure 5). *CHD6* encodes chromodomain helicase DNA binding protein 6 (CHD6), which alters chromatin structure to establish cell type-specific patterns of gene expression.¹¹ CHD6 has multiple domains, including two N-terminal chromodomains, a SWItch/Sucrose Non-Fermentable (SWI2/SNF2)-like ATPase/helicase domain, and a DNA binding do-

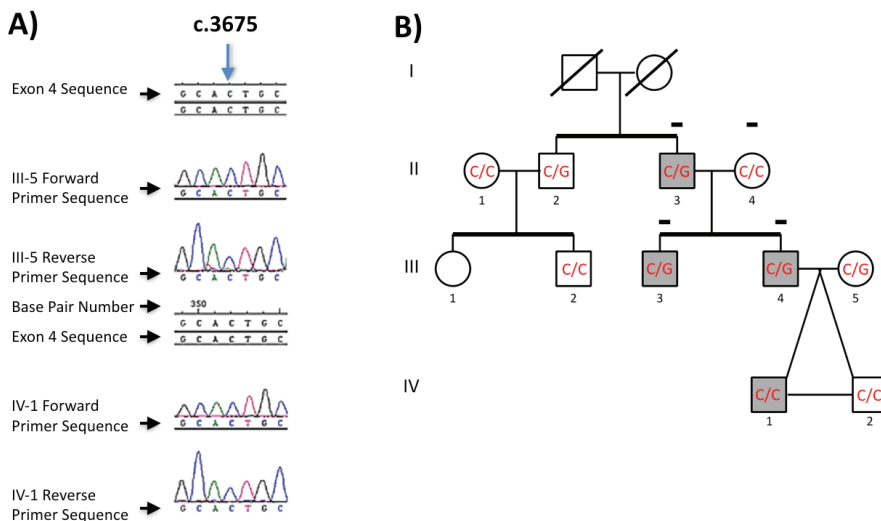


Figure 4: A) DNA sequence data for III-5(top) and IV-2 (bottom) that shows the presence of cytosine at c. 3675 (blue arrow), B) DNA sequence data generated for members of K6821 at c. 3675 (shown in red). Individuals exhibiting the SHFM phenotype are shaded gray.

main that enable the protein to influence gene expression in a cell type-dependent manner.^{20,11,21} In K6821, the heterozygous mutation (c.3675G>C) was predicted to correspond to a change from histidine to glutamine at residue 1225 (p.H1225Q). This change was hypothesized to result in abrogation of regulation of cell type – specific gene expression. As shown in Figure 4, each affected patient except the affected newborn twin (IV-1) is heterozygous for the mutation. However, individuals who do not have the condition also have the mutation (e.g. individuals II-2 and III-5). Finally, the data shown in Figure 4 also shows that IV-1, an affected individual, is homozygous for the allele encoding a normal *CHD6* gene, which suggests that the locus does not play a role in the SHFM phenotype in family K6821.

While the analyses did not elucidate a novel cause of SHFM in family K6821, they do outline a scientific methodology that can be utilized and improved in future studies. Moving forward, it would be ideal to couple these methods along with a genome-wide association study of all patients with SHFM without a known genetic cause. Patients with similar phenotypes could be grouped together, and their exomes could be compared to controls. In summary, these data show that WES followed by confirmatory Sanger sequencing and segregation analysis constitutes a viable method for identifying and evaluating DNA sequence variants in families with a rare genetic disease. Whenever possible, extension of the analysis to newborn affected family members may aid in elucidation of the true etiology of the disease under investigation.

ACKNOWLEDGEMENTS

The authors extend gratitude to Sara Sarasua, PhD for conducting the bioinformatic analysis.

REFERENCES

- ¹Gurrieri, F. and Everman, D. (2013) “Clinical, Genetic, and Molecular Aspects of Split-Hand/foot Malformation: An Update.” *American journal of medical genetics. Part A*. 161A. 11. Pg 2860–72.
- ²Elliott, A. and Evans, J. (2006) “Genotype-Phenotype Correlations in Mapped Split Hand Foot Malformation (SHFM) Patients.” *American journal of medical genetics. Part A*. 140. 13. Pg 1419–27.
- ³Zlotogora, J. (1994) “On the Inheritance of the Split Hand/split Foot Malformation.” *American journal of medical genetics*. 53. 1. Pg 29–32.
- ⁴Birnbaum, R. et al. (2012) “Coding Exons Function as Tissue-Specific Enhancers of Nearby Genes.” *Genome research*. 22. 6. Pg 1059–68.

⁵Tackels-Horne, D. et al. (2001) “Split Hand/split Foot Malformation with Hearing Loss: First Report of Families Linked to the SHFM1 Locus in 7q21.” *Clinical genetics*. 59. 1. Pg 28–36.

⁶Temtam, S. and McKusick, V. (1978) “The Genetics of Hand Malformations.” *Birth defects original article series*. 14. 3, i – xviii. Pg 1–619.

⁷Aken, B. et al. (2016) “Database Update The Ensembl Gene Annotation System.” Pg 1–19.

⁸Aletta, J. and Lewis, S. (1988) “Nerve Growth Factor Regulates Both the Phosphorylation and Steady-State Levels of Microtubule-Associated Protein 1.2 (MAP1.2).” *The Journal of cell ...* 106. Pg 1573–81.

⁹Gould, K. and Nurse, P. (1989) “Tyrosine Phosphorylation of the Fission Yeast *cdc2+* Protein Kinase Regulates Entry into Mitosis.” *Nature*. 342. 6245. Pg 39–45.

¹⁰Lundgren, K. et al. (1991) “*mik1* and *wee1* Cooperate in the Inhibitory Tyrosine Phosphorylation of *cdc2*.” *Cell*. 64. 6. Pg 1111–22.

¹¹Schuster, E. and Stöger, R. (2002) “CHD5 Defines a New Subfamily of Chromodomain-SWI2/SNF2-like Helicases.” *Mammalian genome : official journal of the International Mammalian Genome Society*. 13. 2. Pg 117–19.

¹²Bamshad, M. et al. (2011) “Exome Sequencing as a Tool for Mendelian Disease Gene Discovery.” *Nature reviews. Genetics*. 12. 11. Pg 745–55.

¹³Avila, J., Dominguez, J. and Diaz-nido, J. (1994) “Regulation of Microtubule Dynamics by Microtubule-Associated Protein Expression and Phosphorylation during Neuronal Development.” 25. Pg 13–25.

¹⁴Mandell, J. and Banker, G. (1995) “The Microtubule Cytoskeleton and the Development of Neuronal Polarity.” *Neurobiology of Aging*. 16. 3. Pg 229–37.

¹⁵Brugg, B. and Matus, A. (1988) “PC12 Cells Express Juvenile Microtubule-Associated Proteins during Nerve Growth Factor-Induced Neurite Outgrowth.” *The Journal of cell biology*. 107. Pg 643–50.

¹⁶Diaz-nido, J., Mendez, S. and Avila, J. (1988) “A Casein Kinase II-Related Activity Is Involved in Phosphorylation of Microtubule-Associated Protein MAP-1B during Neuroblastoma Cell Differentiation.” *Journal of Cell Biology*. 106. Pg 2057–65.

¹⁷Letwin, K. et al. (1992) “A Mammalian Dual Specificity Protein Kinase, Nek1, Is Related to the NIMA Cell Cycle Regulator and Highly Expressed in Meiotic Germ Cells.” *The EMBO journal*. 11. 10. Pg 3521–31.

¹⁸Lindberg, R., Quinn, A. and Hunter, T. (1992) “Dual-Specificity Protein Kinases: Will Any Hydroxyl Do?” *Trends in Biochemical Sciences*. 17. 3. Pg 114–19.

¹⁹Patil, M., Pabla, N., Ding, H. and Dong, Z. (2013) “Nek1 Interacts with Ku80 to Assist Chromatin Loading of Replication Factors and S-Phase Progression.” *Cell cycle*. 12. 16. Pg 2608–16.

²⁰Delmas, V., Stokes, D. and Perry, R. (1993) “A Mammalian DNA-Binding Protein That Contains a Chromodomain and an SNF2/SWI2-like Helicase Domain.” *Proceedings of the ...* 90. Pg 2414–18.

²¹Woodage, T. and Basrai, M. (1997) “Characterization of the CHD Family of Proteins.” *Proceedings of the National Academy of Sciences*. 94. Pg 11472–77.

Exercise and Autism Spectrum Disorder: A Neuroscience Perspective

BY ALEX P. KUNZ, MENTOR BROCK JENSEN, PHD
SLIPPERY ROCK UNIVERSITY

Introduction:

Autism spectrum disorder (ASD) is a developmental disorder that permeates throughout all aspects of an individual's life.¹ It is characterized by disordered and/or delayed development of language, restricted or repetitive patterns of behavior, and qualitative impairments in social reciprocity.^{1,2,3,4} ASD affects each individual in a different manner, with different physical and psychological characteristics.^{1,5} Physical and psychological characteristics include self-stimulatory behaviors, infrequent eye contact, lack of verbal initiations, and erratic mood. People with ASD may also experience motor discoordination, hypo- or hypersensitivity to specific sensations, or restrictive interests.^{1,6} Also noted are difficulties in playing imaginatively, participating in social games, sharing/turn taking, holding conversations, and making friends.⁷ While the exact cause is highly debated, these characteristics of ASD can have a major effect on the integration of the individual into the mainstream community.

Evidence suggests that children with ASD may also lack participation in regular physical activity (at least 60 minutes a day) as part of their lifestyle.^{5,6} This may be due to social and behavioral deficits and poor motor coordination. To assist with the prevention in decline of cognitive function, physical activity is needed continuously throughout one's lifespan.⁷ Physical activity may also assist a child in maintaining sustained cognitive tasks that require selective attention and suppressing distractions.⁸ Studies have shown a decrease in aggressive, unproductive, disruptive, stereotyped and self-injurious behaviors through the use of moderate to intensive aerobic exercise.^{4,6} An increase was observed in concentration, attention span, work performance, and on-task behavior.^{4,6} Studies demonstrated that a decrease occurred in specific disruptive behaviors, but not in general to all behaviors; fatigue factor was a major concern for the decrease in these behaviors.^{1,6} The exact effects of moderate to vigorous exercise on the core symptoms of ASD are still being researched and could depend on several factors such as duration, setting, and characteristics of the individual.^{1,9}

The aim of this review is to examine the correlation between exercise and ASD, and explore how exercise may have a positive influence on the social functioning of a child with ASD.

Key Characteristics of Autism:

ASD is diagnosed on the basis of a

two-domain model: social-communication and repetitive and restricted behaviors (RRB) behaviors.^{10,11,12,13,14} Each individual is affected differently, exhibiting diverse core and non-core deficits. Individuals vary greatly in cognitive ability, language proficiency, onset of patterns, and sensory response to environment, with most obvious behavioral symptoms appearing late in the second year of life.^{9,10}

RRBs include rocking, hand flapping, and spinning, with more severe self-harming behaviors presenting only rarely.^{1,6,14} Overall, repetitive behaviors are observed in a large portion of children with ASD, and are distracting to the child with ASD as well as the mainstream community.^{1,15,16}

Impaired modulation of eye contact for social purposes is one of the most critical characteristics used in diagnosis.^{1,3,4,14} Children with ASD show less frequent and shorter fixation on the eyes in various experimental settings.¹⁷ Shorter gaze may interfere with the facilitation of behavioral mimicry. Gaze behavior may serve as an important social function for understanding another individual's mental state and coordinating attention and activities. Processing this direct gaze is not as rapid in individuals with ASD, as compared to typically developing individuals.¹⁸ There is a limitation of visual reciprocity, which suggests neglect in the importance of the eyes as a social information source.¹⁸ Children with ASD show an atypical response to mutual gaze or eye contact; atypical mutual gaze is the individual's inability to mimic expression during passive viewing but the ability to mimic when instructed to do so.¹⁷

Another important diagnostic factor is impaired verbal communication.^{1,13,4} Language development is often atypical and may include an extended period echolalia—repeating other's vocalizations with or without understanding.¹⁹ Children with ASD often have challenges processing auditory or verbal information.¹⁹ Difficulties in perceiving small auditory changes may lead to deficits in communication development, including lack of age-appropriate language growth.¹⁹

Individuals with ASD often prefer solitary activities and spend significantly less time making eye contact; however they do focus longer on a speaker's mouth or body.^{4,19} Face processing has been associated with early potential for sensory processing, and later cognitive function.¹⁹ Deficits in visual motion processing may impact the ability

to process social cues, such as emotional expressions, eye gaze shifts, gesture, and learning and responding to the dynamic environment.⁹ More natural and powerful cues come from dynamic facial expressions than static expressions.²² These behaviors often interfere with positive behavior and provide greater difficulty with integration into the mainstream population.¹

A child with ASD may suffer from a mood that is erratic with poor sitting tolerance and temper tantrums.^{1,3} Social motivation deficit theories suggest children with ASD do not properly recognize or value the pleasures of a social stimulus.²⁰ There may be difficulty in identifying and explaining inappropriate social behaviors, violating social rules, showing increased levels of interpersonal aggression, poor emotional regulation (e.g. tantrums), and reduced positive interaction behaviors.²¹ The individual may poorly distinguish appropriate social behaviors and classify normal behaviors as strange compared to typically developing children.²¹

Social Implication of Autism Spectrum Disorder:

Deficits in social interaction may be expressed in a variety of ways, such as the inability to recognize that others have a point of view, slower reaction times socially, and deficient understanding of emotional facial expressions.^{19,22} Recent models show lower behavioral responsiveness to social incentives, such as competition, which could account for reduced social motivation. This deprives children of crucial social-emotional input thereby hindering the later development of communication skills.²⁰ Children with ASD showed similar benefits from social reward to typically developed children, which can predict the growth of social communication through reward learning.⁵

Autism and the Brain:

Roughly 90% of toddlers with ASD have 5% to 10% – and as high as ~20% – abnormal enlargement in brain volume by the second or third year of life, which will then shift to abnormally slow growth.^{9,12,13,24} The enlarged areas of the brain that are most affected include the frontal, temporal, and parietal lobes, with increased gray matter in the inferior and middle temporal gyri.²⁵ These abnormal developments of the brain are expected to have an impact on auditory development, and visual motion perception, which may lead to conduction

delays that affect functional and structural long-range under-connectivity.^{9,13} Long-range under-connectivity within the brain may affect language acquisition.⁹ During abnormal development, smaller brain size favors long-distance connections, and conversely large brain size favors short-distance connections.¹⁰ Under-connectivity is observed in older children, adolescents, and adults with ASD.²⁶ Abnormal hyperplasia within the brain is also observed in the cerebellar vermis, and cerebral cortex.¹² Children may develop normally for 18-30 months showing subtle symptoms but then fail to progress further or even regress.⁹ Abnormalities in the connections may become more prevalent as the child ages.²⁶ Regression may occur in as many as 20-50% of children with ASD, when observed to have a significant digression in language and nonverbal communication skills.⁹ The timing of the onset of brain development shown through modeling could explain the variety of the disorder's presentations.⁹

Several studies propose that ASD is associated with altered functional and structural connectivity between brain regions, as opposed to deficits in a specific region. Under-connectivity is observed between frontal-parietal or temporolimbic regions, the limbic system and the cerebellum.^{19,25} Also observed are low levels of brain-derived neurotrophic factor (BDNF).^{18,25} BDNF not only protects nerves but also encourages the growth of new neurons.

The following regions have been explored for responsibility for differential encoding and are deficient in ASD: anterior superior temporal sulcus (aSTS), posterior STS (pSTS), intraparietal sulcus (IPS), inferior parietal cortex, fusiform gyrus (FFG), amygdala (AMY), and the dorsal medial prefrontal cortex (dmPFC).^{18,19,22,26} Brain regions involved in social perception and cognition include the AMY, orbitofrontal cortex, and FFG.¹⁹ The FFG, pSTS, dmPFC and AMY deal with direct gaze.¹⁸

The anterior cingulate, dorsolateral prefrontal cortex, the caudate, and the dorsal striatum are engaged during repetitive behavior. The amygdala-hippocampus complex is key in social perception and detects threats and elicits an appropriate behavioral response.¹⁹ The AMY also assists with recognition of social emotions of faces, and the integration of visual and auditory information.^{9,19} Hyperarousal of the AMY during social stimuli could contribute to the social deficits of a child with autism, overall hypoactivation within the AMY is observed in individuals with ASD outside incidents of social stimuli.^{19,22} Difficulty in disengaging from previous tasks may be associated with cognitive inflexibility and is

thought to be caused by poor disconnection of the inhibition network in the frontal-parietal process.¹⁹ Hyperactivation is sometimes observed in the frontal cortex and inferior parietal cortex.¹⁹ The anterior cingulate cortex is an important brain region in social orientation involving emotional self-control, focused problem-solving, and error recognition.¹⁹ The cerebellum is associated with the core symptoms of ASD and cerebellar activation is significantly reduced during attention tasks.¹² Purkinje cell loss is consistently observed in the cerebellum for individuals with ASD, possibly affecting impulsivity.²⁵

Studies have reported a reduction in theory of mind for individuals with ASD.²¹ Theory of mind (ToM) is the ability to attribute mental states to oneself or others and is believed to rely on the medial prefrontal cortex (mPFC), posterior superior temporal sulcus (pSTS), temporoparietal junction (TPJ), and temporal poles.²¹ The medial frontal cortex (MFC) has shown unusual activity in individuals with ASD during tasks in adults and irony comprehension (which is related to ToM) in children.²¹ The MFC contains the mPFC and is responsible for understanding violations of social norms, morality, appropriateness of facial expressions, and person perception and monitoring.²¹ Further, Carter suggests that regional cerebral blood flow is abnormal in the mPFC and that temporal pole activity is found to be abnormal in adults with ASD.²¹ The temporal poles are suggested as a key component in the neural circuitry that is impaired in ASD, and it appears to never deactivate after resting state.²¹

Neural deficiencies have also been observed in the anterior cingulate cortex during monetary reward achievement, and ventral striatum in response to both monetary and social reward.²⁰ These deficits suggest developmental abnormalities in reward circuitry in the frontostriatal and limbic regions, leading to the basis of atypical reward functioning.²⁰ During direct gaze, a lack of context-dependent activity in the STS was observed and showed no difference in activation compared to typically developed individuals and no change between direct gaze and gaze shifts.²⁷ The right anterior insula (AI) – important for regulation of emotion – was not activated, while the left lateral occipital complex (LOC) – important for object identification – was active.²⁷ This shows that some individuals with ASD recruit regions of the brain for processing gaze that a typically developing individual does not.²⁷ TPJ affects individuals with ASD in opposition to typically developing individuals by being activated during averted gaze in ASD versus direct gaze in typical

individuals. Additionally, left dorsolateral prefrontal cortex is diminished during direct gaze similar to the processing of averted gaze in typical individuals.²¹

Brain waves are also affected by ASD including the P300 and N170.^{19,20} Children with ASD generally have a smaller amplitude in the P300 leading to delayed information transmission with no change to diminished response in social reward anticipation.^{19,20} Social deficits correspond with weaker P300 activity in response to reward anticipation, which may have resulted from a deficit in the intervening reward circuitry.²⁰ Individuals with ASD also show longer N170 latencies when distinguishing faces, but normal latencies when distinguishing objects.¹⁹

Children with high functioning ASD have smaller grey matter (GM) volume in the fronto-pallidal region.²⁴ Thinning in the inferior frontal gyrus is correlated with social difficulties while age-appropriate growth does not occur in the GM volume of right inferior parietal lobule, which is important in social perception and posterior cingulate important in social cognition.²⁴ Slowing to stopping of white matter tracts in brain regions associated with social cognition has also been observed.¹² Abnormalities which may result in early overgrowth, followed by slow growth, were predominantly observed in the right inferior parietal lobule and posterior cingulate.²⁴

Characteristics of Exercise:

Higher levels of fitness are associated with higher brain function over a lifespan.²⁴ Physical activity is important for the development of the brain and cognition at childhood.²⁴ During physical activity, a greater level of cognition is required – with more active children outperforming inactive children in cognitive activities.^{24,28} Still it is unclear whether duration of physical activity plays a role in brain changes; children did perform at a higher levels of cognitive function when participating in a 40-minute daily dose of exercise in comparison to a 20-minute dose.²⁹

Physical activity and exercise can provide opportunities for social interaction. Coordinative movements and motor learning play an important role in cognitive function and development, especially attention and the ability to handle visual and spatial information.^{29,30} Improvements in task performance were found after nine months for physically active children; high aerobic capacity helped cognitive function.^{8,28} Low intense physical activity such as resistance training may also improve cognitive function.⁷ Combining physical activity with speech and language stimulation has shown to increase recall, imitation skills, verbal receptive skills,

and expression.³ Physical activity can provide sensory feedback and physiological arousal in a manner similar to self-stimulation.¹ It may also assist with the flexibility of adaptations of behaviors toward specific goals and maintenance of these goals and for teenagers it assists with cognitive development into late adulthood.^{8,29}

Exercise and the Brain:

Single bouts of exercise have been shown to improve cognitive function and neuroplasticity with stimulation after the exercise.^{7,28} Exercise plays a crucial function in hormonal regulation, influencing brain circuits involving various neurotransmitters.²⁸ Physically active individuals show reduced frontal activation after training in areas that were previously active, while increased activation was seen in the frontal, temporal, and parietal lobes.²⁹ Exercise helps assist with enhancing learning ability and memory function, and protects cognitive decline related to aging and neurodegeneration.³¹ Increases in neuroplasticity-related proteins such as BDNF have been observed after bouts of exercise.³²

The frontal cortex plays a major role in cognitive function and development, including attentional selection, working memory, task switching, and inhibitory control.^{8,29} Physically fit children have shown reduced activation of the frontal cortex with a maintenance of attention levels.^{8,29} The frontal region involved in task management are shown to be more hypoactive in physically active children.⁸ The frontal lobe in children is more susceptible to change than that of a mature individual due to the late development of this region.⁷

The right anterior prefrontal region shows a decrease in activation on fMRI scans from pretest to posttest.⁸ Cardiovascular fitness shows higher recruitment of prefrontal and parietal cortex during high cognitive tasks.²⁹ Studies have demonstrated that physical activity positively affects the anterior prefrontal brain function for goal maintenance throughout one's lifetime, which may lead to a child's participation in physical activity presenting more adult-like recruitment of the anterior prefrontal cortex.⁸ Increased brain activation was shown in the frontal and parietal lobes, while a decrease in activation was shown in the anterior cingulate cortex.^{7,29} The anterior cingulate cortex works with the anterior prefrontal cortex in cognitive control.⁸ Coordination training showed decreased activation of the prefrontal region with stretching shown to increase connectivity between the frontal and parietal regions.⁸ This increased connectivity may be beneficial to cognitive function.⁸

Exercise has shown restorative effects

and resurgence of the hippocampus improving cognition, and increased volume of the hippocampus as well as improved connectivity between the hippocampus and the anterior cingulate cortex.^{29,33} Regional central blood flow improved in various brain structures in response to cognitive tasks leading to better task performance.^{7,29,34} Specifically, central blood flow to the hippocampus increased, showing better vascularization of the tissue and increased gray matter.^{7,29,34}

After exercise, cerebral blood flow was shown to improve immediately and longer term improvements were seen in cerebrovasculature.¹⁶ Rats undergoing exercise showed an increased number of Purkinje cells and an increased number of synapses of parallel fibers in the cerebellum and the cerebellar cortex – particularly after thirty consecutive days of training.³⁵ Research has also shown that exercise in adolescent male mice lead to significant changes in the expression of genes associated with synaptic plasticity and signaling pathways.³⁶

Exercise and Autism Spectrum Disorder:

Children with ASD spend less time in moderate physical activity than typically developing children. They do not meet the *Healthy People 2010* standard for adolescents 6-17 years of age, which is sixty-minutes of moderate physical activity every day and twenty-minutes of vigorous activity three days a week.³⁷ One of the limiting factors for not meeting this standard is opportunity, as peers and family play an important role in influencing physical activity regardless of social impairments.^{5,30,37} Avoidance of activity may also be a limiting factor and may be due to difficulties with required skills needed for participation.³⁰ Participation in moderate to vigorous physical activity further declines as the child ages, leading to health concerns associated with inactivity, such as obesity.^{1,37,38,39}

Aerobic exercise has been shown to increase positive behaviors and decrease repetitive behaviors, with the greatest decreases occurring right after exercising.^{6,11,16,39} There is little evidence, however, of the long-term effects of exercise on repetitive motor behaviors. Special education teachers reported better attentiveness and cooperation after physical activity with individuals with ASD.^{1,6,16,39} The physical activity must be of a sufficient length and durability to decrease the repetitive behaviors. However, the length of that change is uncertain. Some have reported an increase in as little as 10 minutes.^{1,6} Intensity has also been shown to play an effect, with more vigorous activity amplifying positive behaviors more than light to moderate physical activity.^{11,39}

More intense aerobic exercise (i.e.,

increase of heart rate above 130 bpm) showed a decrease in negative social aspects of ASD such as hyperactivity and aggression. Moderate activity corresponds with an increase in attention span, on-task behavior, and social play.^{1,4} In rodent models of ASD, treadmill exercise has shown alleviation of aggressive tendency and improved correct decision making in spatial learning memory.³¹ Contingent exercise was also observed to consistently decrease both aggressive actions and comments in individuals with ASD.¹⁵ Following bouts of physical activity; a decrease in negative behaviors was observed along with an increase in positive behaviors such as time on task.^{31,39}

Another form of exercise that has shown a positive impact on ASD is yoga; parents have reported an improvement in the ability of the child to interact with others.³ In addition, there have been observed changes in communication, language, play, and joint attention, as well as eye contact improvement and increased alertness through yoga intervention.³

Discussion:

Some evidence supports the notion that early intervention is beneficial to a child with ASD.¹⁴ Interventions should include behavioral, social, and communication skills associated with the disorder.² More clinicians are recognizing the importance of healthy sleep, nutrition and exercise as part of an early intervention program. The following section is intended to provide some insight and pose several important questions in regards to exercise and its impact on ASD.

Does exercise assist with the activation of the amygdala during rest? Hypoactivation is observed in the amygdala (AMY).¹⁹ The AMY assists with facial recognition, direct gaze, and integration of visual and auditory stimuli.^{9,18,19} When physical activity is combined with speech and language stimulation, improvements are observed in recall, imitation skills, verbal receptive skills, and expression.³ Occupational therapy has similar sensory feedback as moderate physical activity, and has shown an increase in language use and social interaction.¹ When exercised, rats with ASD showed improved decision making in spatial learning memory.³¹

Could exercise impact the activation of the anterior cingulate cortex (ACC) and alter the repetitive behaviors of ASD? Repetitive behaviors may inhibit the individual's ability to integrate into the mainstream community.^{1,16} The ACC is engaged during repetitive behaviors, emotional self-control, and error recognition, and is deficient during monetary reward achievement.^{19,20} Physical activity provides stimulation similar to self-stimulatory behaviors, and decreases the

activation of the ACC.^{17,29} With decreased activation of the ACC, bouts of exercise have shown a reduction in self-stimulatory behaviors, hyperactivity, and aggression with an increase in positive behaviors.^{1,4,6,11,16,39} The restorative effects of exercise have also improved the connectivity between the hippocampus and the ACC.²⁹

Could increasing BDNF encourage the growth of new neurons and assist with improved connectivity? Increased brain size is present in the frontal, temporal, and parietal lobes.²⁵ It has been hypothesized that this increased size may lead to conduction delays, specifically in long-range under-connectivity.¹³ Studies have also reported that ASD is associated with altered functional and structural connectivity.¹⁹ Lower levels of BDNF have been observed.^{18,25} Exercise has been shown to increase BDNF and similar neuroplasticity related proteins.^{8,32} When combined with coordination training, connectivity between the frontal and parietal lobes improves.^{8,32}

*If regular physical activity is maintained, could the hyperactivation of the frontal cortex experienced in children with ASD diminish over time?*¹⁹ With later development in life, the frontal lobe is more susceptible to change in younger individuals.⁷ The frontal cortex is vital in cognitive function and development, reduced activation is observed with exercise.^{8,29} Exercise has also been shown to increase attention span and on task behaviors.¹ A reduced activation after training was observed in the frontal areas that were previously active.²⁹

A limitation of this review was the lack of open access to research that looked at the long-term effects of exercise on ASD. Also, limited research has been performed on the overall effects of exercise on ASD, while specific studies are needed to break down each known individual mechanic of ASD and how it relates to exercise and development. Further studies are required to determine the potential roles of exercise in the development of individuals with ASD including looking at duration, intensity, and type of exercise.

References

1Burns, B., & Ault, R. (2009) "Exercise and Autism Symptoms: A Case Study." *Psi Chi Journal of Undergraduate Research*. 14.2. Pg 43-51.

2Ospina, M., Seida, J. K., Clark, B., Karkhaneh, M., Hartling, L., & Tjosvold, L. (2008) "Behavioral and Developmental Interventions for Autism Spectrum Disorder: A Clinical Systemic Review." *PLoS ONE*. 3.11. Pg e3755.

3Radhakrishna, S., Nagarathna, R., & Nagendra, H. (2010) "Integrated Approach to Yoga Therapy and Autism Spectrum Disorders." *Journal of Ayurveda and Integrative Medicine*. 1.2. Pg 120.

4Borremans, E., Rintala, P., & McCubbin, J. (2010) "Physical Fitness and Physical Activity in Adolescents with Asperger Syndrome: A Comparative Study." *Adapted Physical Activity Quarterly*. 27. Pg 308-320.

5Reid, G. (2005) "Understanding Physical Activity

in Youths with Autism Spectrum Disorders." *Palaestra*. 21.4. Pg 6-7.

6Rosenthal-Malek, A., & Mitchell, S. (1997) "Brief Report: The Effects of Exercise on the Self-Stimulatory Behaviors and Positive Responding of Adolescents with Autism." *Journal of Autism and Developmental Disorders*. 27.2. Pg 202.

7Hötting, K., & Röder, B. (2013) "Beneficial Effects of Physical Exercise on Neuroplasticity and Cognition." *Neuroscience & Biobehavioral Reviews*. 37.9. Pg 2243-2257.

8Chaddock-Heyman, L., Erickson, K. I., Voss, M. W., Knecht, A. M., Pontifex, M. B., & Castelli, D. M. (2013) "The Effects of Physical Activity on Functional MRI Activation Associated with Cognitive Control in Children: A Randomized Controlled Intervention." *Frontiers in Human Neuroscience*. 7.

9Lewis, J. D., & Elman, J. L. (2007) "Growth-Related Neural Reorganization and The Autism Phenotype: A Test of the Hypothesis that Altered Brain Growth Leads to Altered Connectivity." *Developmental Science*. 11.1. Pg 135-155.

10Grzadzinski, R., Huerta, M., & Lord, C. (2013) "DSM-5 and Autism Spectrum Disorders (ASDs): An Opportunity for Identifying ASD Subtypes." *Molecular Autism*. 4.1. Pg 12.

11Petrus, C., Adamson, S. R., Block, L., Einarson, S. J., Sharifnejad, M., & Harris, S. R. (2008) "Effects of Exercise Interventions on Stereotypic Behaviors in Children with Autism Spectrum Disorder." *Physiotherapy Canada*. 60.2. Pg 134-145.

12Won, H., Mah, W., & Kim, E. (2013) "Autism Spectrum Disorder Causes, Mechanisms, and Treatments: Focus on Neuronal Synapses." *Frontiers in Molecular Neuroscience*. 6. Pg 19.

13Lewis, J. D., Theilmann, R. J., Townsend, J., & Evans, A. C. (2013) "Network Efficiency in Autism Spectrum Disorder and its Relation to Brain Overgrowth." *Frontiers in Human Neuroscience*. 7. Pg 845.

14Alabdali, A., Al-Ayadhi, L., & El-Ansary, A. (2014) "Association of Social and Cognitive Impairment and Biomarkers in Autism Spectrum Disorders." *Journal of Neuroinflammation*. 11.1. Pg 4.

15Luce, S. C., Delquadri, J., & Hall, R. V. (1980) "Contingent Exercise: A Mild but Powerful Procedure for Suppressing Inappropriate Verbal and Aggressive Behavior." *Journal of Applied Behavior Analysis*. 13.4. Pg 583-594.

16Anderson-Hanley, C., Tureck, K., & Schneiderman, R. (2011) "Autism and Exergaming: Effects on Repetitive Behaviors and Cognition." *Psychology Research and Behavior Management*. 4. Pg 129-137.

17Senju, A. (2013) "Atypical Development of Spontaneous Social Cognition in Autism Spectrum Disorders." *Brain and Development*. 35.2. Pg 96-101.

18Georgescu, A., Kuzmanovic, B., Schilbach, L., Tepest, R., Kulbida, R., Bente, G. (2013) "Neural Correlates of 'Social Gaze' Processing in High-Functioning Autism under Systematic Variation of Gaze Duration." *NeuroImage: Clinical*. 3. Pg 340-351.

19Belger, A., Carpenter, K. L., Yucel, G. H., Cleary, K. M., & Donkers, F. C. (2011) "The Neural Circuitry of Autism." *Neurotoxicity Research*. 20.3. Pg 201-214.

20Kohls, G., Peltzer, J., Shulte-Ruther, M., Kamp-Becker, I., Remschmidt, H., Herpertz-Dahlman, B., et al., (2011) "Atypical Brain Response to Reward Cues in Autism as Revealed by Event-Related Potentials." *Journal of Autism and Developmental Disorders*. 41. Pg 1523-1533.

21Carter, E. J., Williams, D. L., Minshew, N. J., & Lehman, J. F. (2012) "Is He Being Bad? Social and Language Brain Networks during Social Judgment in Children with Autism." *PLoS ONE*. 7.10. Pg e47241.

22Sato, W., Toichi, M., Uono, S., & Kochiyama, T. (2012) "Impaired Social Brain Network for Processing Dynamic Facial Expressions in Autism Spectrum Disorders." *BMC Neuroscience*. 13. Pg 1-17.

23Sacrey, L. R., Germani, T., Bryson, S. E., & Zwaigenbaum, L. (2014) "Reaching and Grasping in Autism Spectrum Disorder: A Review of Recent Literature." *Frontiers in Neurology*. 5. Pg 1-12.

24Cheng, Y., Chou, K., Fan, Y., & Lin, C. (2011) "ANS: Aberrant Neurodevelopment of the Social Cognition Network in Adolescents with Autism Spectrum Disorders." *PLoS ONE*. 6.4. Pg e18905.

25Viola, S., & Maino, D. (2009) "Brain Anatomy, Electrophysiology and Visual Function/Perception in Children within the Autism Spectrum Disorder." *Optometry & Vision Development*. 40.3. Pg 157-163.

26Pelphrey, K. A., Shultz, S., Hudac, C. M., & Wyk, B. C. (2011) "Research Review: Constraining Heterogeneity: The Social Brain and its Development in Autism Spectrum Disorder." *Journal of Child Psychology and Psychiatry*. 52.6. Pg 631-644.

27Pitskel, N. B., Bolling, D. Z., Hudac, C. M., Lantz, S. D., Minshew, N. J., & Wyk, B. C. (2011) "Brain Mechanisms for Processing Direct and Averted Gaze in Individuals with Autism." *Journal of Autism and Developmental Disorders*. 41.12. Pg 1686-1693.

28Ando, S., Hatamoto, Y., Sudo, M., Kiyonaga, A., Tanaka, H., & Higaki, Y. (2013) "The Effects of Exercise Under Hypoxia on Cognitive Function." *PLoS ONE*. 8.5. Pg e63630.

29Voelcker-Rehage, C., & Niemann, C. (2013) "Structural and Functional Brain Changes Related to Different Types of Physical Activity across the Life Span." *Neuroscience & Biobehavioral Reviews*. 37.9. Pg 2268-2295.

30Yazdani, S., Yee, C. T., & Chung, P. J. (2013) "Factors Predicting Physical Activity Among Children with Special Needs." *Preventing Chronic Disease*. 10.

31Seo, T., Cho, H., Shin, M., Kim, C., Ji, E., & Baek, S. (2013) "Treadmill Exercise Improves Behavioral Outcomes and Spatial Learning Memory through Up-Regulation of Reelin Signaling Pathway in Autistic Rats." *Journal of Exercise Rehabilitation*. 9.2. Pg 220-229.

32Pareja-Galeano, H., Briochio, T., Sanchis-Gomar, F., Montal, A., Jovani, C., Martinez-Costa, C., et al., (2013) "Impact of Exercise Training on Neuroplasticity-Related Growth Factors in Adolescents." *Musculoskeletal and Neuronal Interactions*. 13.3. Pg 368-371.

33Rodgers, S. P., Trevino, M., Zawaski, J. A., Gaber, M. W., & Leasure, J. L. (2013) "Neurogenesis, Exercise, and Cognitive Late Effects of Pediatric Radiotherapy." *Neural Plasticity*. Pg 1-12.

34Viboolvorakul, S., & Patumraj, S. (2014) "Exercise Training Could Improve Age-Related Changes in Cerebral Blood Flow and Capillary Vascularity through the Up Regulation of VEGF and eNOS." *BioMed Research International*. 2014. Pg 1-12.

35Garcia, P. C., Real, C. C., Ferreira, A. F., Alouche, S. R., Britto, L. R., & Pires, R. S. (2012) "Different Protocols of Physical Exercise Produce Different Effects on Synaptic and Structural Proteins in Motor Areas of the Rat Brain." *Brain Research*. 1456. Pg 36-48.

36Abel, J. L., & Rissman, E. F. (2013) "Running-Induced Epigenetic and Gene Expression Changes in the Adolescent Brain." *International Journal of Developmental Neuroscience*. 31.6. Pg 382-390.

37Obrusnikova, I., & Cavalier, A. (2011) "Perceived Barriers and Facilitators of Participation in After-School Physical Activity by Children with Autism Spectrum Disorders." *Journal of Developmental and Physical Disabilities*. 23.3. Pg 195-211.

38Bandini, L. G., Gleason, J., Curtin, C., Lividini, K., Anderson, S. E., & Cermak, S. A. (2013) "Comparison of Physical Activity between Children with Autism Spectrum Disorders and Typically Developing Children." *Autism*. 17.1. Pg 44-54.

39Macdonald, M., Esposito, P., & Ulrich, D. (2011) "The Physical Activity Patterns of Children with Autism." *BMC Research Notes*. 4.1. Pg 422.

40American Psychiatric Association. (2013) *Diagnostic and Statistical Manual of Mental Disorders (5th ed.)*. American Psychiatric Association.

The Role of Impulsivity in the Effectiveness of Restrictive Eating Behaviors

BY SARAH NACHIMSON

STERN COLLEGE FOR WOMEN, YESHIVA UNIVERSITY

Abstract

Restrictive diets are a typical and well-recognized solution to the obesity pandemic. This literature review discusses six studies which assessed the effectiveness of restrictive eating behavior as it pertains to weight loss and the mediating role of impulsivity. The paper first looks at studies that examined the overall efficacy of restrictive eating. Results from the studies suggest that the rigidity aspect of the dieting process accounts for its overall ineffectiveness. The latter portion of the paper evaluates research which determined the effect of impulsivity on dieting. Results from the studies discussed indicate that, by lowering impulsivity levels, dieters might increase their likelihood of succeeding. Future research might therefore be aimed at assessing methods of improving impulse control in dieters.

Keywords: diet, impulsivity, weight loss, food restriction

The effectiveness of dieting as a weight management strategy is a widely discussed topic, both in everyday conversation and formal literature. Anecdotal evidence and empirical research have suggested that dieting is not the perfect solution for shedding unwanted pounds, as many people follow various dieting programs for extended periods of time with little or no success. The prevalence of obesity in the world has been deemed a major health problem, and therefore dieting, a potential solution to the problem, demands consideration. In an analysis for the Global Burdens of Disease Study, researchers demonstrated that the rates of being overweight or obese rose 27.5% in adults and 47.1% in children from 1980 to 2013.¹ The issue is certainly an urgent one, yet the study related that there have not been any national success stories reported in the last 33 years regarding the obesity pandemic. While it may seem that dieting is ineffective at combating obesity, research shows that there may be a specific factor that makes some people more successful in their diets than others. The following studies have examined the efficacy of rigid dieting in healthy weight management and the role of impulse control in contributing to dieting success.

In their 2013 study, Goldstein, Katterman, and Lowe noted the ineffectiveness of dieting and concluded that dieting does not lessen caloric intake from food.² By using self-report, the experimenters distinguished between participants dieting to lose weight, participants trying to avoid weight gain, and participants not dieting at all. They also evaluated the extent to which participants were using food restriction for the purpose of weight control and grouped subjects according to restraint status as unrestrained and restrained eaters. Perhaps as an oversight, the study does not specifically explain the difference between the two groups. In effect, the four groups studied

were those dieting to lose weight, those dieting to avoid weight gain, restrained non-dieters, and unrestrained non-dieters. The researchers hypothesized that unrestrained non-dieters would have the highest caloric intake and current weight loss dieters would have the lowest caloric intake. They used 24-hour food records to evaluate food consumption. Results from the study did not confirm the researchers' hypothesis, as they found no significant difference between the four groups studied in terms of their total caloric intake. Interestingly, researchers did find that unrestrained non-dieters consumed more calories from drinks than those who were more conscious about their caloric intake. However, when caloric intake from food was analyzed separately from caloric intake from beverages, there was no significant difference between the groups. Thus, results from this study demonstrated that dieting does not reduce caloric consumption from food and might therefore be ineffective at combating obesity.²

Neumark-Sztainer et al. (2006) similarly demonstrated the inefficiency of restrictive eating behavior, namely its tendency to predict both weight gain and unhealthy eating patterns.³ Their study examined adolescents who were dieting or engaging in other weight-controlling behaviors. It aimed to assess whether these behaviors contributed to weight gain, overweight status, disordered eating habits, and occurrences of eating disorders five years later. At the beginning of a five-year period, participants were asked how often they had dieted during the previous year, and dieters were distinguished from non-dieters based on their responses. Additionally, participants identified their weight-controlling behaviors. Healthy behaviors, such as eating fewer sweets and more fruits and vegetables, were distinguished from unhealthy behaviors, such as fasting and taking diet pills. Participants'

body mass index (BMI) was also measured. When BMI was measured five years later, results indicated that dieting and weight-control behaviors did not lead to decreased BMI or overweight status. Instead, dieting and related behaviors were linked to weight gain, and five years later, dieters were almost twice as likely as non-dieters to be classified as overweight. The role of unhealthy weight-control behaviors in predicting overweight status was particularly evident. Dieting was also associated with an increased risk of developing eating disorders.³

The above studies both emphasized the ineffectiveness of dieting in healthy weight management. People who were dieting did not actually eat less than those who were not dieting.² Moreover, not only was dieting associated with weight gain, but it was found to induce eating disorder patterns as well.³ Clearly, dieting does not produce a desired outcome for many people. What the studies left unexplained, however, is whether there is a particular property inherent in the dieting process which may contribute to its failure.

Through a study that looked at rigid and flexible control in dieting, Meule, Westenhofers, and Kubler provided insight as to why dieting may fail in many circumstances.⁴ Researchers expected dieting success to decrease with rigid control, or inflexibility in dieting, and increase with flexible control, or a more balanced approach to food consumption. Their procedure involved collecting data from participants via an online survey that measured rigid and flexible eating control strategies, food cravings, and perceived success in dieting. A long version of the flexible and rigid control scale was employed in the study to increase its reliability. Results confirmed that rigid control was a significant predictor of unsuccessful dieting, whereas flexible control was related to successful dieting. Additionally, only rigid control was associated with food cravings.⁴

Meule et al. (2011) highlighted the

function of inflexibility in predicting an unwanted outcome for dieters.⁴ Rigid control has indeed been found to inversely predict dieting success, but rigidity is often a central component of popular weight loss diets. However, despite the seemingly ineffective nature of such diets, certain characteristics of dieters may increase or decrease their likelihood of achieving weight goals through restrictive eating. According to Logan, Schachar, and Tannock, “people who are impulsive have trouble inhibiting action, whereas people who are not impulsive find it easier to do so.”⁵ Impulsive individuals might therefore experience more difficulty in following restrictive diets than individuals who are not impulsive.

Research has demonstrated how the role of subject impulsiveness mediates the effects of the dieting process. The stop-signal task, which requires participants to deliberately inhibit an anticipated response, is often used to measure impulsivity. Participants are instructed to respond as quickly as possible to a go task and to refrain from responding upon hearing an occasional stop signal. A participant’s ability to inhibit his/her response after hearing the stop signal indicates good impulse control.⁵

Nederkoorn, Jansen, Mulkens, and Jansen predicted that children who were most impulsive would lose the least weight in a weight loss treatment program.⁶ Impulsivity in 26 obese children was measured with the stop-signal task before a behavioral treatment designed to decrease overweight status was administered. Results demonstrated that impulsivity was directly correlated with the percentage at which a child was overweight: children who were more impulsive were more overweight. Additionally, children with higher levels of impulsivity lost less weight from the treatment than children with lower impulsivity. Children who were most overweight lost the least weight in the study, as losing weight seemed more difficult for them.⁶

Nederkoorn et al. elucidated an important connection between impulsivity and dieting success, but their study contained several limiting factors. One apparent factor which was lacking in the study was a description of the specific nature of the behavioral treatment administered to the obese children.⁶ Because the level of restraint in dieting may play a role in its successfulness, it would be interesting to note whether the treatment involved highly restrictive eating rules, flexible rules, or a combination of both types of guidelines.⁴ The study addressed the role of impulsivity in determining dieting success, but it failed to address whether an interaction effect exists between rigidity of treatment and the impulsivity of dieters.

Additionally, because the study examined the effects of impulsivity on dieting in obese children, results may be limited in their generalizability to an adult population.⁴

In another study, Jansen et al. also hypothesized that dieting success is related to impulsivity.⁷ They expected that, when tempted, eaters with high restraint who were more impulsive would eat more than high-restrained eaters who were less impulsive. They did not expect to see a difference in food intake between participants with low restraint who had high impulsivity and participants with low restraint who had low impulsivity. Food intake was the dependent variable in this study, and it was measured during a fake taste test. Experimenters measured food intake in participants on three different occasions after participants either ingested a preload consisting of two milkshakes, smelled high-caloric foods, or completed a questionnaire that was utilized as a time filler. Participants were told that the goal of the study was to examine how tasting and smelling affect reaction time and that they should eat as much of the food presented to them as they wanted. They were also instructed to complete a questionnaire about taste perception as well as the stop-signal task. The restraint scale was completed by participants instead of the stop-signal task only after their final taste test in order to prevent them from exhibiting more restraint in future trials. Results showed that an interaction between restraint and impulsivity is related to overeating. Restraint alone did not determine overeating, but participants who were high in restraint ate more if their impulsivity was also high.⁷

Koningsbruggen, Stroebe, and Aarts further confirmed that impulsiveness is linked to unsuccessful dieting in restrained eaters more than unrestrained eaters.⁸ Researchers predicted that for unrestrained eaters, level of impulsiveness would either not affect dieting success or affect it to a lesser extent. They also predicted that dieting would be more difficult for restrained eaters who were more impulsive. Participants in the study completed questionnaires that assessed their trait impulsiveness, dietary restraint, and perceived dieting success. Results confirmed the research hypothesis, as experimenters found that higher impulsivity was connected to lower success in dieting for restrained eaters but not for unrestrained eaters.⁸

While Jansen et al. (2009) and Koningsbruggen et al. (2013) both examined the interaction between restraint and impulsivity levels in eating habits, the results of the former may be more reliable than the those of the latter.^{7,8} Food intake in the first experiment was measured with a fake taste test, and thus there was an objective measure

for the amount of consumption. However, the procedure for the second experiment involved a questionnaire, which evaluated perceived dieting success. The use of a questionnaire rather than an observable experiment to measure food consumption allowed for participant subjectivity, as the definition of dieting success for one participant might be different than the definition of success for another. Therefore, researchers from the first study could be more certain that their results were accurate and not based on the subjective experiences of participants.

Dieting can be an arduous process, and furthermore, following a rigid diet can cause people to overeat rather than eat less. Research has demonstrated that although strict dieting does not predict weight loss, people who have lower impulsivity levels can avoid the negative consequences associated with rigid eating behaviors. Hence, the role of impulsivity is central to the effectiveness of restrictive eating plans, as it can predict dieting success. These research findings lend themselves to practical application in terms of their ability to delineate ways to improve dieting strategies. In order to facilitate success in the dieting process, it would presumably be helpful if future studies were to assess methods of decreasing impulsivity levels in dieters. In this way, findings of past research could make a significant contribution to increasing the effectiveness of food restrictive behaviors.

References

- ¹Ng, M., Fleming, T., Robinson, M., Thomson, B., Graetz, N., Margono, C., ... Gakidou, E. (2013). Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*, 384, 766–781.
- ²Goldstein, S.P., Katterman, S.N., & Lowe, M.R. (2013). Relationship of dieting and restrained eating to self-reported caloric intake in female college freshmen. *Eating Behaviors*, 14, 237–240.
- ³Neumark-Sztainer, D., Wall, M., Guo, J. Story, M., Haines, J., & Eisenberg, M. (2006). Obesity, disordered eating, and eating disorders in a longitudinal study of adolescents: How do dieters fare 5 years later? *Journal of the American Dietetic Association*, 106, 559–568.
- ⁴Meule, A., Westenhof, J., & Kubler, A. (2011). Food cravings mediate the relationship between rigid, but not flexible control of eating behavior and dieting success. *Appetite*, 57, 582–584.
- ⁵Logan, G.D., Schachar, R.J., & Tannock, R. (1997). Impulsivity and inhibitory control. *Psychological Science*, 8, 60–64.
- ⁶Nederkoorn, C., Jansen, E., Mulkens, S., & Jansen, A. (2007). Impulsivity predicts treatment outcome in obese children. *Behaviour Research and Therapy*, 45, 1071–1075.
- ⁷Jansen, A., Nederkoorn, C., van Baak, L., Keirse, C., Guernieri, R., & Havermans, R. (2009). High-restrained eaters only overeat when they are also impulsive. *Behaviour Research and Therapy*, 47, 105–110.
- ⁸van Koningsbruggen, G.M., Stroebe, W., & Aarts, H. (2013). Successful restrained eating and trait impulsiveness. *Appetite*, 60, 81–84.

Female War Correspondents and their Effect on Female Military Positions

By KARLIN ANDERSEN

WHITWORTH UNIVERSITY

Women began supporting military efforts in official and trained units during World War II as administrative assistants, secretaries, or nurses, organized through the Women's Army Corps (WAC). Their positions; however, were temporary and limited in resources and status. This inequality led to a 68-year journey to bring women into equal status with men in the armed forces through a string of legislation and alterations in military rules and regulations. Immediate credit belongs to the women who served in WWII and beyond; however, female war correspondents also aided in altering public opinion toward the notion of women working in and around the military through their coverage of the WAC, and eventually from within units. During the WAC's progress to gain increased status within the Army, these correspondents worked to attain access to battlefields and soldiers, earning respect from the military and their own colleagues as far back as World War I. Through the work of these civilian female correspondents in tangent with the WAC's progress, the concept of women on and around the battlefield became more feasible and widely accepted, especially after the Vietnam War. The involvement of female war correspondents did not independently lead to the advances for women in the military, but their presence on or near the frontlines and their persistence to be allowed closer to battles and to speak with officials contributed to women demanding greater access and equal treatment.

The WAC began their work in 1943 as the first non-auxiliary organized group of women trained for the military.¹ Training began at Fort Des Moines in Iowa, with their officers eventually deployed to England, North Africa, and other areas of minimal fighting.² The male labor shortages during WWII prompted the government to turn to 2.5 million previously unemployed women to fill jobs as secretaries, mail clerks, radio operators, repairwomen, mechanics, and medical lab technicians, with peak WAC participation reaching just under 100,000.^{1,2} Women were seen as a valuable asset in order to free up men to fill combat roles, but the presence of women in military positions was debated and "met with resistance in public opinion" as well as in Congress. Both political leaders and the public felt "a mix of both disdain and protectiveness toward women," which

led to the consensus that "women were not to go into combat."³

Due to the WAC's origin as a temporary support system for WWII, the corps was scheduled to dismantle six months after the war ended; however, the women's substantial contribution to the military's effort in Europe prompted General Dwight D. Eisenhower to propose legislation in 1946 to permanently integrate the WAC into the Army and the Organized Reserve Corps. With the WAC becoming a permanent fixture of the Army in 1948, women were given validation for their work in WWII and allowed a larger platform to strive toward increased roles within the Army.² Their change in status was also a sign of a continued shift in public and Army opinion from the years during WWII until 1948. Soldiers and leadership within the Army may have been swayed by witnessing the WAC's contributions and assistance in Europe. However, domestic voters would not have been able to witness those same efforts and experiences without a middle party to translate the battlefield scenes to America—this is where the female war correspondent stepped in to cover the WAC. Often assigned to women because the WAC was considered less dangerous and of more interest to women than the male soldiers' duties, all aspects of the WAC—from training in the U.S. to service in Europe—were photographed and written about throughout WWII. The coverage these female correspondents provided allowed American citizens to understand how the WAC was contributing to WWII, and why the eventual 1948 bill to establish them within the Army mattered.

The WAC's permanent existence allowed the organization to continue recruiting and training volunteers in the years between WWII and the Vietnam War, leading to a tripling in enlistment compared to WWII numbers, and playing an important role in the WAC's ability to fulfill the need for supplies and assistance in the Vietnam War. The WAC's increased presence in the Vietnam War led to increased equality in rankings, benefits, and retirement for officers. Women could be promoted above the rank of lieutenant colonel, and women in the WAC were able to serve in the Army National Guard in 1967.² By the end of the Vietnam War, women were authorized to participate in Reserve Officers' Training Corps (ROTC) programs, which created

a pool of women willing and able to enter the Army service upon graduation. While on paper, women achieved training and benefits during the Vietnam War, their presence in Vietnam was limited to a peak of 109 women overseas, while a total of 2.7 million men served.⁴ The WAC's progress, despite their limited numbers, can be attributed to the way in which they successfully completed the jobs they held in Vietnam and at home as administration, personnel, logistics, signal, transportation, intelligence, and other operational specialties.⁴ Due to the limited number of women directly interacting with soldiers and higher military officials, the large number of female reporters in Vietnam assisted in balancing the gender inequality in Vietnam and the concept of women on the battlefield through their work. Unlike female reporters who moved with units, women officers only traveled near combat zones and aided military personnel. This work demonstrated their emotional and physical strength in a way that had not been seen in past wars.

In theory, WAC volunteers and nurses in Vietnam should have been safe in their restricted noncombat zones. However, because battles in Vietnam employed guerilla warfare, and thus often occurred in villages, on highways, or as surprise attacks, both women reporters and members of the WACs were often caught in battle areas. The Army recognized the experiences women had serving and completing tasks under the stress of surprise attacks, increasing training opportunities after the war—but assignments were still restricted to "safe" areas.⁵ Although in 1973, women were allowed to train as pilots and later admitted to the majority of officer or enlisted specialty positions, they were excluded from entering combat or training for combat positions due to the 1977 Combat Exclusion Policy.¹ With the 1978 disbandment of WAC and full integration of women into the military, women were allowed to train with men. However, a policy created ten years later, the Standard Risk Rule, established an assignment rating system to assess risk of "direct exposure to hostile fire or capture," which resulted in further restrictions on women from areas deemed too dangerous or risky.²

The stigma that women could not handle the physical or emotional stress of combat missions persisted into the 1980s. During

this time, the WAC struggled with position equality while female war correspondents also attempted to reach battlefields. The presence of female reporters among soldiers, along with the challenges female reporters overcame from both the military and within their own profession helped to prove women's capability to survive wars, breaking down social barriers for women and further propelling the concept of women serving in the military forward. Although the WAC did not exist until WWII, female correspondents began their largest push for foreign assignments in WWI by applying for accreditation to reach the battlefield and report on the war firsthand. However, they were told to cover the "woman's angle," or stories behind the front lines about civilians.⁶ Those limitations caused reporter Corra Harris to believe that, "Being banned from the front because of sex provided to be the biggest obstacle for women journalists."⁶ The restricted experience female reporters gained in WWI instilled the desire to cover all news— both foreign and domestic—which was carried forward to the next generation of reporters in WWII.

WWII brought the same hurdle of accreditation, which required a receipt of placement, passing security clearances and receiving approved travel documents before going overseas to report. Put in place to keep track of overseas reporters, accreditation also served as an easy way to approve male reporters for travel whilst keeping women out of Europe.⁶ In WWII, women were refused based on the burden the military would face to provide food, clothing, housing, and protection. Once those logistics were resolved, the military and domestic government offices that handled accreditation expressed concerns over the need for "special facilities" or toilets and sleeping areas to be created for women, which further stalled accreditation efforts. Eventually, roughly 120 to 140 women reporters overcame the government obstacles to receive accreditation.^{3,6}

While successful in their efforts to achieve access to Europe, their stories were often restricted to "reporting on the activities of female military personnel," although this allowed "Americans to see that they were safe and leading normal lives" while in Europe.⁶ The reports and photos of American women working within the military illustrated to the nation that women can operate within the military. Toni Frissell, famous for her photographs in *Vogue* and *Harper's Bazaar*, photographed women training for WAC and under review.

Her photos "fit into [the] media campaign to counter negative public perception of women in uniform."⁷ Similar to the WAC,

reporters stepped in to fill positions vacated by men during war. While the roles were never directly related in creation or duties, their coexistence helped to prove the importance of each: female correspondents documented the WAC's achievements, and the WACs provided a platform for female correspondents to prove their reporting abilities. Female reporters often acted as stand-ins for WAC positions on the front lines, and they gave the military exposure to how women may act in a military setting.

Margaret Bourke-White, a photographer for Life Magazine, was the first female photographer accredited in WWII. Through her work in 1941 photographing the German bombing of the Russian Kremlin, she served as an example of how women acted beyond the base and how they could work with soldiers.³ Bourke-White could attend "early dawn briefings [and] go on practice fights," and she also marched with General Patton's army into Buchenwald, one of the first liberated German concentration camps. However, she was not permitted to do "the one thing that really counted—go on an actual combat mission."³ Her survival of bombing raids and her photographs of some of the war's worst tragedies gave her professional merits, but she was unable to photograph battles for personal safety reasons. The limitations placed upon her did not diminish the effect of her photos as they served as a testament to women's ability to work in a war zone and to continue to function in the face of war tragedies. Bourke-White's presence among soldiers and in active war zones aided women in the military who held positions across Europe and Northern Africa, but could not fully enter battle areas themselves. Reporters in WWII served as the proxy for female soldiers by allowing male soldiers and commanders to experience women's abilities both on base and in dealing with the horrors of war.

Clare Boothe Luce served that role when she wrote on-location articles during WWII, including articles on the military effort. In 1941, she wrote a profile for Life Magazine on General Douglas MacArthur when he was promoted to Commanding General of all U.S. Armed Forces in the Far East. Luce's piece stepped outside of the common career profile to focus on Gen. MacArthur's choices during the war, and how she believed the war could progress. Within a short article, she could examine how the Allied Powers viewed the war and propose, "that in Tokyo this chapter head would cause many a fan-toothed naval jaw to drop and many a military bandy leg to buckle."⁸ Through her specific analysis of military strategy, Luce stepped outside of the traditional female role and asserted her knowledge of the

current events of the time and progression of the war. Her work provided valuable background on Gen. MacArthur's military career along with how he could contribute to a second front in the east if needed. Her knowledge of Gen. MacArthur and military strategy demonstrated within that time an ability for women to travel to combat zones, comprehend the war, and digest the information to make astute observations. While Luce's article may not have altered tactics, her presence and astute predictions of military strategy provide evidence for her and other women's capabilities within military operations.

Unlike women in the WAC, who were kept away from battle, female reporters in WWII were allowed deeper into Europe, and traveled with soldiers into Germany and Russia. Female reporters also continued to push for job equality in domestic newsrooms, while the WAC struggled with decreased volunteers due to the Army's original plan to dismantle the. While women reporters' progress in war reporting did not make large strides until Vietnam, their continued efforts in newsrooms around the country allowed them to make greater advancements in Vietnam and place themselves deeper in combat than in previous wars.

During the Vietnam War, the persistence of female reporters to enter battle zones and travel with soldiers continued to indicate to the military that women were interested in and capable of experiencing battle conditions, completing their assignments, and surviving. Their commitment began with getting to Vietnam: editors of papers and radio and television stations usually did not hire female foreign correspondents; thus, any interested female reporter had to convince an editor of both their credentials and their willingness to travel to Saigon. Some reporters even went to Vietnam without a job, hoping to write freelance and eventually be picked up by a paper or radio station. This included Denby Fawcett, who paid for her own plane ticket before being picked up by the *Honolulu Advertiser's* in 1966.⁹

The *Honolulu Advertiser's* editors restricted Fawcett to the cities and tasked her to write "color stories on Saigon and its environs."⁹ Her movements were further limited by military officials who "did not like the idea of male reporters getting killed" and "were against female soldiers in combat."⁹ One official told Fawcett she reminded him of his daughter and was "even more horrified at the thought of a woman reporter getting shot."⁹ Fawcett reflected on this by asserting that the commander "would never say to a male reporter, 'You remind me of my son.'"⁹ Eventually, she was able to talk her way past officials and checkpoints to begin

traveling with soldiers on helicopters and into firefights. Once she asked to walk point squad, “the most dangerous and exposed unit,” which travels at the forefront of the action, open to booby traps, mines, and ambushes.⁹ The experience Fawcett gained aided her in obtaining access to more battles, better interviews, and longer stays with units.⁹ However, she still faced opposition from higher commanders, including General William Westmoreland, commander of U.S. armed forces in Vietnam, who—after encountering Fawcett on a base tour—attempted to remove all female reporters from Vietnam. He was met with resistance from a group of female reporters who lobbied the Pentagon to continue their reporting work, but according to Fawcett, “the incident showed how tenuous our hard-earned privileges were.”⁹ In Vietnam, Fawcett could move away from the relative safety of Saigon and her editors, which demonstrated that she was able to march with soldiers while other female reporters, aid workers, and members of the WAC were kept behind battle lines. Other women in Vietnam were held back by government and military restrictions, but reporters who often reported to domestic publications, which could not directly control their moves in Vietnam, were able to move more freely throughout South Vietnam once they received clearance or found a unit willing to take them on. The ease of embedment allowed female reporters to undergo traditionally masculine experiences and write from the direct perspective of someone in the field. That placement also gave the military, both soldiers and leadership, an understanding of how women may operate while under the pressures of war. During the two decades of the Vietnam War, these efforts pushed public and private government opinion toward understanding how women could withstand war.

Another key figure in this push for more freedom was Edith Lederer. When she requested to travel outside of Saigon in 1972, her editors at *Associated Press* (AP) placed restrictions on her. The AP foreign editor did not think “women could stand up to the demands of the world’s backwaters and battle zones,” but AP’s president called Lederer with an idea for a promotion to cover the extraction of U.S. troops.¹⁰ Despite the company president allowing female reporters in Vietnam, he made it clear to Lederer she was not to report on fighting. With this direction, Lederer felt she had “to prove that [she] was as good as AP’s male war correspondents.”¹⁰ Lederer’s desire is illustrative of women’s push for professional equality with their male counterparts, and the ability to report on and work within the same areas as male reporters. Just as women

serving in the military received the “safe” assignments, female correspondents were often told to stay within city limits and not report on battles. While WAC officers and volunteers were restricted by placement, correspondents made progress throughout the war as they talked their way past security checks, proved their merit through their articles, and took the place of male reporters who had been reassigned to other domestic or foreign assignments. Lederer and Fawcett’s repeated attempts—both failures and successes—to gain access to battles which put them in proximity with military commanders and front line soldiers. Fawcett and Lederer’s expressed desire to “cover exactly the same kinds of stories as [their] male colleagues” helped build the mindset that female reporters did not want to be treated differently than male reporters. This mentality was translated onto battlefields where women were expected to care and protect themselves just as any soldier would.

One of Gen. Westmoreland’s fears rested in his belief that female reporters “might inconvenience and endanger soldiers who would rush to protect us in firefights.”⁹ Fawcett believes Westmoreland also “worried that the women correspondents might collapse emotionally when faced with the horrors of combat,” an idea which Fawcett calls “unwarranted.” Fawcett recounts “in combat it was usually every man and woman for him—or herself.” While remembering an attack on a village while traveling with a battalion of Marines, she recalls, “I was essentially alone when the shooting started.” Kate Webb, a reporter for United Press International in Vietnam, proved women’s capability to withstand the worst of war when she survived capture by North Vietnamese soldiers for 23 days. Taken prisoner along a highway during a battle in 1971, Webb and five other reporters marched through the night with minimal food and water while suffering from sicknesses and violent interrogation by their captors. Webb described her time as a prisoner as “the gray limbo,” and stated that she was “not among the living or the dead of the war, but trapped in a gray twilight with no links to the living world.”¹¹ Webb reflected that the fact that she survived gave hope to other families with missing soldiers or aid workers. Other women, both reporters and aid workers, were also captured in Vietnam, but Webb’s publicized story displayed how women could operate successfully during war even as a prisoner.

Those stories helped to fill in the gaps when the military limited women serving within their ranks. Through their action on the frontlines of WWI, WWII, and the Vietnam War, these women were not

forgotten by the military or the public, especially when these women pushed their editors to publish their stories. Frissell’s photographs in WWII of the WAC, along with the work of other photographers after her, provided an important glimpse into the WAC’s contributions to the collective war effort. Each female journalist and photojournalist throughout these three wars contributed evidence—in both their work and their presence—that women can enter a war zone and continue their jobs while working alongside others. This contributed to a shift in how women in the military were viewed, as they were then seen as valuable assets—not only in supporting roles, but in the military as well. This led to continued growth in positions and opportunities throughout the 2000s. As historian Janie Blankenship noted, “it was [the] women stationed in war zones throughout U.S. military history that paved the way for those in Afghanistan and Iraq.”¹² This refers to the WAC, but also extends to female reporters who, by the Vietnam War era, lived in war zones for months or years to photograph and report on the conflict for domestic news organizations. The uninterrupted presence and continual efforts to seed themselves further into combat areas demonstrated how female correspondents supported and aided the development of the military’s acceptance of women in general and within the armed forces.

References

- ¹[N.A.] “Timeline of women reporting war.” *No Job For A Woman*. [N.D.]
- ²Burgess, J.G. “History of army women.” *United States Army Women’s Museum*. [N.D.]
- ³Yellin, E. (2004) *Our Mother’s War*. Pg 47, 111, 116.
- ⁴[N.A.] “Women in the army comparative statistics by war.” *United States Army Women’s Museum*. [N.D.]
- ⁵[N.A.] (2012) “Highlights of women’s contributions to the United States Army.” *United States Army Women’s Museum*.

The Royal Navy's Blockade System 1793-1805: A Tactical Paradox

BY PATRICK BRASZAK

WILFRID LAURIER UNIVERSITY

Throughout the French Revolutionary and Napoleonic Wars, Great Britain relied heavily upon its maritime forces, known collectively as the Royal Navy. The Royal Navy's blockade of the European continent during this period thwarted Napoleon Bonaparte's ambitions for the invasion of Great Britain, cut France off from its overseas resources and facilitated England's rapid imperial and economic growth.¹ The blockade system, however, only became a viable strategy once Great Britain had achieved naval supremacy over France following the Battle of Trafalgar. Prior to this pivotal moment, the Royal Navy's blockade system had put its ships at a tactical disadvantage against enemy vessels. This fundamental flaw, masked by the broader successes of the blockade system, was only overcome by the skill of the Royal Navy's ablest admirals and captains.

There were numerous difficulties that blockading squadrons had to contend with while at sea. Ports enjoy the luxury of concentrated land and naval defences, making attack impractical. Blockading squadrons had to cope with bad weather, which could drive them from their positions and allow enemy fleets opportunities to escape port — or enter port. The blockade system was ruinous to the ships, hard on the morale of seamen, and incredibly expensive. Blockades also robbed the Royal Navy of public support throughout the Wars because they inhibited significant fleet actions from occurring.²

The Royal Navy's greatest strength was its prowess in battle. Its men were better trained than those of their adversaries; at the time of the Revolutionary War, British gun crews could fire three times faster than the French crews.³ This allowed the British to capture far more French and Spanish ships than they lost throughout the Wars, despite the inferior designs of the Royal Navy's vessels. The French and Spanish men-of-war had broader beams and deeper draughts, which meant that they were larger, sturdier, and capable of carrying far greater armaments.⁴ The Spaniards' monstrous men-of-war that stood four decks tall, such as the *Santísima Trinidad* and the *Santa Anna*, were exemplars of the technical advantage which Britain's adversaries enjoyed.⁵ That these two vessels sunk following the Battle of Trafalgar shows the mastery of the

Royal Navy at open-water engagements.⁶

The Royal Navy's blockade system counteracted their advantage at sea because it discouraged French and Spanish ships from leaving port, meaning decisive open-water conflicts could not occur.⁷ The Royal Navy kept a constant presence outside its enemy's ports to prevent hostile expeditions against Great Britain. By doing so, the British contained their adversaries but allowed them to maintain a fleet in being: not active, but always ready to become so. The British would commit the same tactical error during the First World War — despite being under blockade, Germany's *Hochseeflotte* posed a continuous threat to Britain's merchant navy, until it was thoroughly diminished at the Battle of Jutland.⁸ Regardless of the historic period, Britain's naval enemies were predominantly safe under a blockade system, and furthermore posed a constant threat to British domestic and foreign interests.⁹

Many of the Royal Navy's most prominent figures recognized these tactical deficiencies, and thus despised the blockade system. This included Richard Howe, who, as Admiral of the fleet, held the Royal Navy's highest rank. During his correspondence with Lord Chatham, Howe argued that the Royal Navy should keep its fleets anchored within the relative safety of friendly ports, with only frigates acting as sentinels to watch the enemy's coast for movement.¹⁰ This method ultimately encouraged open-water naval battles. It would be easier for enemy fleets to put to sea without British men-of-war anchored off of their ports — but their movements would be relayed back to the Royal Navy, who would then move its fleets to intercept. Howe reasoned that:

(...) the two contending fleets might then engage on something like equal terms, as to their state and condition, each of them fresh from their respective ports; whereas a blockading squadron, keeping the sea for months without being relieved, and exposed to all kinds of weather, ought not to be considered on a par with an enemy of equal force fresh from a port, and still less in a condition to follow them, perhaps to a foreign station (...)¹¹

Howe's confidence in the Royal Navy's prowess was well-founded, for he himself had been involved in a victorious action

against a larger fleet early in the French Revolutionary War.¹² On May 2, 1794, Admiral Howe (then the commander-in-chief of the Channel fleet) had put to sea from Port St. Helens with twenty-two men-of-war and six frigates in search of a French convoy that he had heard was coming back from North America and the West Indies.¹³ Howe's intelligence was validated when, on May 19, an American vessel out of Brest reported that a French fleet composed of twenty-four men-of-war and ten frigates had left two days earlier to protect their valuable homeward-bound convoy.¹⁴ Outnumbered by six ships, Howe nonetheless set off, and on May 25, he destroyed two French corvettes that had mistaken his ships for their own convoy. Knowing that the French were close, Howe steered northwards until on May 28, several of his advance frigates spotted the opposing fleet far to the south-east. Howe gave chase and was rewarded when the *Revolutionnaire*, a three-deck warship, struck its flag by the end of this first day.¹⁵

Howe spent the beginning of the following day enduring shots directed at his van; the French fleet would pull away, then haul their wind and concentrate their fire upon the leading British ships before again fleeing.¹⁶ He decided at noon to give the signal to break through the French line, but after this maneuver was completed, the British found themselves too dispersed to be able to concentrate on any portion of the segmented French fleet. May 30 and 31 were spent recovering from the confusion, which was prolonged due to particularly foggy conditions. On the morning of June 1, however, the French admiral decided to haul to the wind and form a line with his twenty-six sail (some ships had joined the French fleet from nearby ports while others had become separated).¹⁷

The action between the two fleets ensued at nine in the morning. After a little more than one hour of furious fighting, the French admiral decided to flee northward where he again formed a line with roughly a dozen of his ships. Noticing this, ten more French ships bore away to join their comrades; Admiral Howe's fleet was too dispersed, and had sustained too much damage to their masts and rigging to prevent the escape. Seven French ships, however, were too crippled to flee, and were thus abandoned to be seized by the British.¹⁸ The British vessels

could be repaired, but the French had lost a significant portion of their naval force by abandoning them to Howe's fleet.

Admiral Howe's gallant victory, dubbed "the Glorious First of June," exemplified the Royal Navy's prowess and revealed why many of its most prominent figures abhorred a blockade system designed to avoid similar battles. One such man was Vice-Admiral Lord Collingwood, who had been present on that occasion sailing under Admiral Howe as (then) Captain of the *Barfleur*.¹⁹ Unbeknownst to him at the time, Collingwood was destined to spend the remainder of his life fighting the French. From the beginning of the French Revolutionary War in 1793 to his death in 1810, Collingwood spent all but one year at sea. It was a fitting culmination to a perilous fifty-year career — the seasoned Vice-Admiral had served forty-four years abroad in active employment by the end of his life.²⁰ As a dedicated servant of country and cause, Collingwood's disapproval of the blockade system lent significant credence against its tactical efficiency.

Collingwood's disdain of blockading grew steadily throughout the French Revolutionary and Napoleonic Wars. In April 1796, Collingwood joined the blockade off Toulon and quickly realized how monotonous the entire affair would be. In a letter written from the *Excellent* on May 11, 1796 to J. E. Blackett, Collingwood described the situation this way: "(...) [the French] cannot move a ship without our seeing them, which must be very mortifying to them; but we have the mortification also to see their merchant-vessels going along shore, and cannot molest them."²¹ French land batteries and heavy moveable artillery units prevented Collingwood from getting close enough to the shore to disrupt French shipping in any meaningful way. The role of his fleet, with its powerful men-of-war, had been largely reduced to merely observing the French lying in port.

As the blockade of Toulon went on, this monotony grew to be a very real danger to the British ships because of France's exploits on land. Collingwood wrote another letter to Blackett on September 25 of that year explaining that the French had closed the ports at Leghorn and Genoa to the Royal Navy, which meant that his fleet would have difficulty securing the supplies they needed to maintain their health. On December 5, Collingwood expressed his fears that his fleet could possibly be trapped in Porto Ferrajo if they were not cautious, for it was the Royal Navy's only friendly port left in the region following the evacuation of Corsica. Collingwood insisted that his fleet could beat the enemy's if they met at sea, but the threat of the enemy waiting instead to corner the

British once they had to resupply was too great to continue the blockade.²²

While Collingwood's experiences blockading Toulon exemplified the dangers of waiting outside an enemy's port while war waged on inland, his experiences off Cadiz illustrated the difficulties of bringing a blockaded enemy to battle. Collingwood's frustrations with Cadiz were apparent on January 26, 1798: he had been goading the enemy's fleet of thirty men-of-war by cruising with just six of his own, yet was consistently denied release from boredom by the enemy's ardent refusal to sail from their port. Collingwood had received intelligence indicating that the Cadiz fleet planned to join twenty-six other line-of-battle ships from Carthage and Toulon. He could do little to prevent this dangerous possibility because Collingwood had become convinced that the Cadiz fleet would only move when they were certain they could avoid battle at sea.²³ The soldiers aboard his ships were growing so restless from the lack of action that Collingwood had to resort to creative distractions to quell their agitated states.

We have lately been making musical instruments, and now have a very good band. Every moonlight night the sailors dance; and there seems as much mirth and festivity as if we were in Wapping itself. One night the rats had destroyed the bagpipes we made, by eating up the bellows; but they suffer for it, for in revenge we have made traps of all constructions, and have declared a war of extermination against them.²⁴

While the enemy (meaning the French and Spanish ships, not the rats), whose numbers were vastly superior to his own, would not come out to meet him, Collingwood was also apprehensive of bringing the battle to their port for fear of their gun-boats. Mounted with heavy cannons and propelled by oars and sail, these tiny vessels were able to avoid most of the fire that line-of-battle ships and frigates could offer them. The gun-boats were also far more maneuverable in light winds and shallow waters, which robbed the Royal Navy's larger vessels of much of their speed and dexterity.²⁵ A stalemate was thus inevitable, for neither side was willing to initiate any sort of action.

Such stalemates often gave the tactical advantage to the fleet anchored in port, as was made evident by Collingwood's dismayed letters to Blackett on December 9, 1798 and August 18, 1799. Dejected, Collingwood wrote on the ninth that the French could set sail from their ports almost at will, noting that a fleet had joined the Spaniards off Carthage.²⁶ The latter correspondence was as follows:

In all reasonable expectation, the French fleet ought not to have escaped us; and I had always hopes of our coming up with them, until we sailed into Port Mahon, which is a very narrow harbour, from which you cannot get out without great difficulty. There we remained, until the enemy had got so far the start, that it was not possible to come up with them. We arrived at Brest the day after them, and finding them snug, came [to Torbay]; - at all which there has been great lamentation in the fleet.²⁷

Despite the considerable efforts that had gone into containing the enemy fleets, they had managed to escape and travel through the Straits of Gibraltar, and then a considerable distance northwards along the coast of Spain and France.

It appears the only measure that could have been taken within the blockade system to prevent the escape of such a combined fleet was to have attacked them in their own ports. Lord Horatio Nelson, who was arguably the Royal Navy's most influential commander of the period, was the only British officer who would achieve any considerable success at doing so in the French Revolutionary and Napoleonic Wars. He had attacked and defeated an enemy's port on two separate occasions: the first outside Alexandria, and the second at Copenhagen.²⁸

Nelson's encounter outside Alexandria in August of 1798 had been the culmination of a long and arduous pursuit. On April 30, 1798, Earl St. Vincent had detached Nelson from his duties before Cadiz so that he could watch French forces at Toulon.²⁹ Earlier that year, Nelson had mentioned in passing to his wife that "Buonaparte is gone back to Italy, where 80,000 men are embarking for some Expedition."³⁰ He would go on to pursue this expedition for four months. While sailing for Toulon, Nelson's squadron captured the French corvette *la Pierre*, which divulged that

The French General, Buonaparte, arrived at Toulon ten days ago, to command the secret Expedition preparing to sail from that port (...) It was not, however, generally believed that Buonaparte would embark; but no one knows to what place the Armament is destined. Nineteen sail of the line are in the harbour, and fifteen are apparently fitted for sea: yet it is said, that only six are to sail with the transports now ready, and that 12,000 men are now embarked.³¹

Nelson's unique vigour was made apparent by a private note that he added in his letter to his patron, Lord St. Vincent, which read "Be assured I will fight the

French fleet the moment I can find them; until then adieu.”³² It was this determination to not only fulfill his duty, but to destroy his enemy that separated Nelson from his colleagues and allowed him to succeed in a blockade system predisposed to stalemates.

When Nelson finally arrived at Toulon on June 4 after a month marked by a violent squall, there were no French ships to be seen. With neither instructions nor intelligence on what to do next, Nelson decided to sail for Corsica, following the northwest wind that he assumed the French had also taken. After he again found nothing, Nelson steered towards Naples, and was finally rewarded on June 20 with information that the British consul at Malta had surrendered to the French. While on course to Malta, Nelson was rewarded with a second instance of luck, when a Genoese brig reported on June 22 that the French had sailed four days prior from Malta at the head of a northwest gale. Deducing correctly that the enemy was aiming for Egypt, Nelson took off in that direction.³³

In an unfortunate occurrence of mistiming, Nelson arrived at Alexandria on June 29 only to find a governor shocked that he was chasing a French fleet. Puzzled that his enemy had not arrived, the admiral set off on a course that took him past several islands on the way to the port at Syracuse in Sicily, which ironically made him miss the French fleet a second time. On July 25, Nelson would again sail for Egypt only after receiving assurances that the French fleet had gone to neither the Archipelago, the Adriatic, nor the Mediterranean. Nelson finally got his ultimate reward when on August 1, he again came within sight of Alexandria, and at four in the afternoon received a signal that the approach to the city was full of French ships.³⁴

Nelson’s fleet had found sixteen ships anchored in Aboukir Bay. Their defensive position gave them the decided advantage against any attack, with only their artillery to focus their attention on whilst the British would have to navigate the narrow harbour under heavy fire. Nevertheless, at a quarter past six in the morning on August 2, the first British ships approached their enemy, receiving fire from the van ships, coastal batteries and gun boats. Nelson ordered his fleet to double the French line, and within twelve minutes, the French *Guerrier* had lost its masts. The *Conquerant* was dismasted just ten minutes later.³⁵

Losses were severe for both the British and the French. Nelson anchored his *Vanguard* on the other side of the enemy’s line, and within just a few minutes, “every man stationed at the first six guns in the forepart of the *Vanguard*’s deck, were all either killed or wounded, and one gun in particular was repeatedly cleared.”³⁶ So

ferocious was the action that at nine in the morning, the French flagship *L’Orient* caught fire. It fought on for thirty-seven minutes until the flames finally ignited its magazines, causing the “*L’Orient* to [blow] up with a crashing sound that deafened all around her. The tremulous motion, felt to the very bottom of each ship, was like that of an earthquake.”³⁷ The massive explosion silenced both sides for a few minutes, until the French began shooting again at a frenetic pace, spurred on by their anger and grief.³⁸

The fighting laboured on until three o’clock the next morning, when both sides stopped, exhausted. This second break lasted only an hour, until the *Alexander* and *Majestic* resumed their attacks on the *Tonnant*, *Guillaume Tell*, *Genereux*, and *Timoleon*. More British ships steadily joined the fray, and the two sides continuously bombarded each other until Rear-Admiral Villeneuve cut the cable of the *Guillaume Tell* at eleven in the morning and made off with the *Genereux* and two frigates. The British ships, too crippled to follow, resolved themselves to securing the spoils of their victory. They had captured a total of eleven men-of-war and two frigates, more than in any other naval victory of the Napoleonic Wars.³⁹

The victory at Aboukir Bay brought praise to both the Royal Navy and Lord Nelson.⁴⁰ He had managed to assault and defeat a fleet of heavily-defended ships in harbor, with just one of his own having to strike its flag after running into the rocky bottom of a particularly shallow portion of the bay.⁴¹ Nelson would prove his merit again almost three years later when on April 2, 1801, he led an attack on Copenhagen harbour, whose defences were even greater than those he had encountered off Alexandria.⁴² Copenhagen had been defended by

(...) six sail of the line, eleven floating batteries, mounting from 26 twenty-four pounders to 18 eighteen pounders, one bomb ship and schooner gun vessel. These were supported by the Crown islands, mounting 88 pieces of cannon, four sail of the line moored across the mouth of the harbour, and batteries on the island of Amok.⁴³

With just twelve ships, Nelson plunged headlong into a hail of Danish fire. Three of his ships, the *Bellona*, *Russel*, and *Polyphemus*, were immediately taken out of the action, running aground in the shallow and narrow water.⁴⁴ The *Bellona* and *Russel* did not quit however. They were “within range of shot, and continued to fire with much spirit upon such of the Enemy’s ships as they could reach.”⁴⁵ The perseverance of the *Bellona* and *Russel* could have been attributed to Nelson’s leadership; he trusted the capabilities of all the men sailing under him, and always encouraged them to take initiative if they came upon an

opportunity that deviated from his battle orders.⁴⁶

The battle began five minutes after ten in the morning. Nelson maneuvered his own ship, the *Elephant*, into the centre of the action, with one British ship astern and three ahead.⁴⁷ Neither side had gained any decided advantage by one in the afternoon, at which point Sir Hyde Parker, aboard the *London*, gave the signal for the British ships to withdraw from the fight. Nelson ignored this signal for a moment, and then gave the order to simply acknowledge it rather than repeat it, and then carried on his bombardment of the Danish defences.⁴⁸ After four hours of desperate fighting, in which droves of Danish reinforcements strove to keep the British at bay, Nelson was finally able to subdue the defences and secure a truce that ended Denmark’s part in the Second Coalition.⁴⁹

Lord Nelson was a remarkably successful naval officer, but his prowess alone could not make up for the flaws of the Royal Navy’s blockade system. An entire tactical system cannot rely on the ingenuity of a single man. Lord Nelson had been an unusually aggressive, ambitious and insightful individual. Nelson’s peers, although distinguished and proven in their own rights, simply could not fight in the same way he did. The frustrated efforts of a Captain Peter Puget (who had been employed in the blockade of Brest) to organize an attack on the moored French ships characterized the difficulties that men without Nelson’s abilities faced when trying to work within the blockade system.

Puget had first proposed his idea of using fire ships — vessels deliberately set afire and steered into enemy formations — to destroy the French fleet at Brest on June 23, 1804, in a letter to Admiral Cornwallis. Enclosed in his letter had been an incredibly detailed report which outlined the specific number and class of ships that would be needed for the operation, along with the requisite number of men, equipment and sailing conditions. Puget asked for:

(...) ten brigs from 100 to 130 tons each, and three sloops from 50 to 60 tons each (...) They should be supplied with grapnels and chains, as also with two fast-rowing six-oared Deal boats (...) Two captains, ten lieutenants, ten master’s mates, ten midshipmen, with a hundred seamen, would be sufficient to conduct the brigs; three lieutenants, three midshipmen, and eighteen seamen, the sloops; in all 156 persons, the whole to be volunteers.⁵⁰

With this force, Puget argued that:

The best time for making the attempt

to destroy the enemy's fleet in Brest appears to be in the last quarter of the flood tide, on a starlight night, with the wind blowing a commanding breeze from the N.W., or, indeed, further to the westward, for the boats could always secure their retreat with the assistance of the ebb tide.⁵¹

Puget also explained the path that the attacking squadron should take to reach the enemy, the order by which the fire ships should attach themselves to the enemy, and the exact steps that should be followed in securing a retreat.

In going in, the fire-brigs should keep close under the north shore, on the principle that the nearer the vessels are carried to the enemy's batteries, the effect of red-hot shot would be lessened, for they would pass through both sides, instead of lodging (...)

(...) form the brigs in a loose, compact line. When the leading vessels had arrived off the battery of Portzic, they should, with the first division, haul up for the Southern part of the French fleet; the second division for the northern part (...). The fire-brigs should be on board the enemy, if possible, a quarter of an hour before high water, for it would be of essential benefit in assisting them to maintain their station (...). It would also be the means of more effectively spreading the flames, as with the flood tide and wind (...).⁵²

His meticulous efforts were rewarded when, on August 17, Cornwallis gave his consent for the plan to proceed.⁵³

Per Cornwallis' letter, the eight fire ships that were to attack Brest should have prepared to rendezvous with their squadron by August 29. Just three days before this date however, Lord Melville of the Admiralty wrote to Cornwallis to inform him of the plan's first delay. Inspectors had examined the fire vessels and had proposed that some of the ships should be substituted, meaning that all of them would have to wait before setting sail. Melville would write again on September 5 to announce that the vessels were finally ready, but a shift in winds to the westward two days later slowed the arrival of the fire ships further.⁵⁴

The result of these delays was Cornwallis' orders on September 29 and 30, which instructed the lieutenants of the gun-vessels and cutter that were to comprise the retinue of fire ships to return to their previous stations, effectively cancelling Puget's plan.⁵⁵ In another letter on October 16 addressed to Lord Melville, Cornwallis explained that

The particular service had unavoidably

been delayed considerably beyond the time, and Captain Puget had observed that the enemy had changed their position as well as some gun-vessels. The unfavourable opinion of the sea officers of the [Naval] Board did also dampen the ardour of some.⁵⁶

Captain Puget, unlike Nelson, had not been free to attack off his own initiative when the opportunity had been available. His proposition had been vulnerable to unforeseen delays, unfavourable weather conditions, and a non-compliant enemy.

The Royal Navy's blockade system certainly did not facilitate engagements with the enemy. Admiral Howe had noticed the negative effects that this inactivity could have on public opinion early in the French Revolutionary War. British newspapers had repeatedly slandered Howe in 1793 because of his failure to bring the enemy to battle. Poor weather had forced the admiral to return to port with the Channel fleet on four separate occasions since he had received his secret instructions on July 1, 1793; the last occasion had forced Howe to retire to Torbay for nearly four months until the necessary repairs were made. Blockades were effective at protecting trade and damaging the enemy's commercial interests, but the public wanted to see their war taxes spent on "the destruction of the enemy."⁵⁷

The blockade system's negative impacts on public opinion were exemplified by the Prime Minister William Pitt's motion on March 15, 1804, for an inquiry into the country's naval administration. This inquiry was meant to prove that Lord St. Vincent, who was First Lord of the Admiralty at the time, had insufficiently prepared the Royal Navy for the defence of Great Britain from invasion.⁵⁸ In particular, Pitt had accused the Royal Navy of failing to construct the types of ships that it needed to defend Great Britain, of exerting less effort in fighting the enemy and preparing itself than it had previously in the French Revolutionary War, and in failing to secure enough contracts to build the necessary number of ships of war.⁵⁹

This motion showed that there was fear outside the Royal Navy that Great Britain was in danger, despite its control of the English Channel. In a response to Pitt's accusations, several naval officers proved that his motion had been premised upon false principles. Pitt had believed that gun-vessels were needed to defend Great Britain's shores from invasion when, in reality, the larger vessels that already constituted the Royal Navy's permanent fleet were preferable to execute that task. Frigates and larger vessels could mount a greater proportion of guns, were far more durable than any gun-vessel could be, were immune to being boarded by the French

gun-vessels, and required a smaller amount of men to oppose Napoleon's massive invasion flotilla.⁶⁰ Napoleon had amassed 130,000 men, over one thousand boats, and 8,500 horses, all spread over seven ports: the major one being Boulogne, with smaller harbours at Vimereaux, Ambleteuse, and Étaples as well as North Sea ports at Calais, Dunkirk, and Ostend.⁶¹ England simply did not have enough men to arm a fleet of gun-vessels large enough to counter Napoleon's; according to the official answer to Pitt's motion, half of the Royal Navy's seamen and marines at the time would have been needed to cover Boulogne alone.⁶²

Pitt also argued that gun-vessels were more effective in shallow waters, but the naval officers retorted that there was a far greater proportion of deep water between Great Britain and the European continent that needed to be defended. The capability of gun-vessels to operate near coasts was also of little benefit offensively, for land batteries and heavy moveable artillery units commonly defended enemy ports.⁶³

Had the Royal Navy been engaging the French fleets in open waters, rather than allowing them to wait in port, public opinion would likely have been very different. Lord Nelson's most famous victory, the Battle of Trafalgar, by no means ended the threat posed by enemy fleets or Napoleon's invasion ambitions, but it nevertheless convinced the British public that they had been saved from danger.⁶⁴ The Brest fleet remained following the battle, and a new one would be raised at Toulon.⁶⁵ Furthermore, after subjugating Russia, Austria and Prussia in 1807, Napoleon would once more turn his attention to Great Britain. Russia, Denmark, Norway and Sweden all threatened to oppose the Royal Navy in 1809, yet the magnitude of the Royal Navy's victory at Trafalgar gave the British confidence that they were invulnerable to any further threats that could come over the sea.⁶⁶

The Battle of Trafalgar was fought on October 21, 1805. A combined French and Spanish fleet of thirty-three men-of-war had set sail from Cadiz on October 19, only to be immediately pursued by twenty-seven of their British counterparts. The chase lasted two full nights. At daybreak on October 21, the combined fleet noticed that the British had separated into two columns; the French and Spanish concaved their line in response, so that their rear and van could fire upon the sides of the approaching British ships whilst the middle shot at the British front.⁶⁷ The British ships, utterly vulnerable, occasionally fired shots to cover their advance with smoke.⁶⁸

Admiral Collingwood, who was leading

one of the columns aboard the *Royal Sovereign*, broke through the enemy line at noon. A melee of flying splinters, crashing masts and smashing hulls commenced. Collingwood gave the first ship he passed, the *Santa Anna*, (...) a broadside and a half into her stern, tearing it down, and killing and wounding 400 of her men; then, with her helm hard a-starboard, she ranged up alongside so closely that the lower yards of the two vessels were locked together. The Spanish Admiral, having seen that it was the intention of the *Royal Sovereign* to engage to leeward, had collected all his strength on the starboard; and such was the weight of the *Santa Anna's* metal, that her first broadside made the *Sovereign* heel two streaks out of the water.⁶⁹

The fury of the moment prevented anyone aboard the *Royal Sovereign* from calculating exactly how long their ship had been fighting alone, but after at least twenty minutes, the other British ships cut through the line as well.⁷⁰

Nelson had led his column in the *Victory* towards the enemy's van in what would come to be known as "the Nelson touch," feinting to mask his true intention of piercing its centre where the flagship under Vice-Admiral Villeneuve was positioned.⁷¹ As he closed in on his adversary, the enemy ships surrounding Villeneuve fired probing shots to gauge Nelson's distance. Then finally "(...) a shot passed through the *Victory's* main-top-gallant-sail; the hole in which being perceived by the Enemy, they immediately opened their broadsides, supporting an awful and tremendous fire."⁷² Twenty men were killed and thirty wounded before the *Victory* had even fired a purposeful shot in response.⁷³

Nelson's flagship finally crashed through the enemy's line at twenty minutes past noon. The *Victory* ran aboard the *Redoubtable*, the two becoming locked together in a mess of rigging and cannon fire. The *Temeraire* came to Nelson's aid and ran itself aboard the other side of the *Redoubtable*, but not before the *La Fouguenx* had managed to run alongside it. The four men-of-war fought on, entangled in the centre of the raging melee. It was during this frenzy that Nelson received his mortal injury, as a French marksman shot him through the left breast.⁷⁴

The fighting lasted three full hours until Admiral Gravina finally broke off with ten ships towards Cadiz. Minor skirmishes only followed, when five of the enemy ships decided to return in an effort to save their comrades. These were beaten off, and the British squadron claimed

(...) nineteen ships of the line

(of which two are first-rates, the *Santissima Trinidad*, and the *Santa Anna*), with three flag-officers, viz. Admiral Villeneuve, the Commander-in-chief; Don Ignacio Maria d'Alava, Vice-admiral; and Rear-admiral Don Baltazar Hidalgo Cisneros.⁷⁵

In the end, three of the enemy's vessels managed to flee back into Cadiz, but these had been reduced to wrecks during the battle. Fourteen ships from the combined fleet were either burnt, sunk, or crashed onto shore, and 20,000 men were taken prisoner.⁷⁶ When news of this victory reached Great Britain, the populace broke out into celebration.⁷⁷ Their Royal Navy had managed to win a decisive naval engagement against a blockaded enemy. The Battle of Trafalgar was not, however, a product of the blockade system that had been employed up until this point; a unique set of circumstances, along with a conscious effort on behalf of the British to remain hidden led the combined fleet to its decision to set sail and leave Cadiz.

Trafalgar was more a product of Howe's alternative plan to the blockade system.⁷⁸ The battle originated not when the combined fleet left port on October 19, but rather on August 20, 1805. Collingwood had been sailing off Cadiz with three men-of-war, one frigate, and a bomber when he encountered thirty-six French and Spanish vessels. The combined fleet was sailing for the port when they noticed Collingwood's small force and sent sixteen ships after him. In a display of skillful seamanship, the British squadron managed to outmaneuver and elude the pursuing enemies as it fled towards the Straits of Gibraltar. After the French and Spanish gave up the chase, Collingwood reversed and resumed his station of Cadiz, sending one of his ships out onto the horizon to make fake signals, as if to a larger force of British ships so that the combined fleets would be intimidated into remaining in port.⁷⁹

When news of Collingwood's encounter spread, Nelson was immediately dispatched to provide his assistance. While on route, Nelson wrote to Collingwood asking that "(...) no salute may take place, but also that no colours be hoisted: for it is as well not to proclaim to the enemy every ship which may join the fleet."⁸⁰ Repeating Howe's ideas, Nelson sent another letter to Collingwood on October 10 that read "(...) if the weather is fine (...) and we are in sight, [the combined fleet] will never move; and should it turn bad, we may be forced into the Mediterranean, and thus leave them at liberty to go to the westward (...)"⁸¹ He understood that the French and Spanish needed to be lured out of Cadiz as quickly as possible, and that the most effective way to do so was to mask the numbers of the growing British squadron.

The combined fleet could not remain

in port at Cadiz for long. Napoleon had prepared Brest, Rochefort, and Ferrol with enough stores for a force of its size, but not Cadiz. The combined fleet had only arrived there after Sir Robert Calder had forced it south. Collingwood was aware of this, and on October 6 wrote to Nelson that the combined fleet had become completely isolated from any supplies out of France.⁸² It was under these circumstances that the fresh British squadron tempted the combined fleet with an avenue to escape, knowing that they were desperate and could be intercepted. The Battle of Trafalgar occurred because the British had actively sought to bring the enemy out from Cadiz's harbour, rather than containing it through blockading methods.

The destruction of the combined fleet at the Battle of Trafalgar altered Napoleon's plans for Great Britain, and thus had significant strategic consequences for the Royal Navy. The emperor's primary focus following the battle was no longer invasion, but rather to defeat Great Britain by ruining its commerce with the Continental System.⁸³ Napoleon's Berlin Decree of November 21, 1806, had placed all the British islands under blockade, and had forbade British vessels from entering French ports. He followed this order up on December 17, 1807, with the Milan Decree, which declared any ships that submitted to searches by the Royal Navy to be denationalized.⁸⁴

Napoleon had enacted his Decrees without having any fleets capable of enforcing them. Great Britain, by contrast, had issued Orders in Council on May 16, 1806, and November 11, 1807, which effectively ended France's maritime trade. The first Order in Council had declared the entire European coast from Brest to the Elbe under blockade, while the three issued on November 11 did the same to every port that belonged to a nation or colony hostile to Great Britain. The Royal Navy was now charged with maintaining a paper blockade; it did not have the resources to watch every port as closely as had been required prior to the Battle of Trafalgar, but its unquestioned dominance of the seas meant that it did not have to. The Royal Navy was able to employ its fleets in such a manner as to make the approaches to the European continent sufficiently dangerous as to dismay any merchant vessels from attempting to defy their blockade. They "(...) controlled the approaches from the Atlantic to all the northern continental ports; and at Gibraltar those to the Mediterranean."⁸⁵ The Royal Navy's blockade system was now its greatest asset, acting as a deterrent which facilitated the growth of Great Britain's commerce and industrial power while simultaneously draining France of its resources and straining Napoleon's ties with his continental allies.⁸⁶

From the years 1793 to 1805, the Royal

Navy had attempted to contain enemy naval forces through an onerous blockade system which required its ships to maintain a constant presence outside hostile ports. This approach to naval warfare was tactically inefficient because it allowed enemy fleets to maintain fleets in being, it did not allow the Royal Navy to take advantage of its superior fighting forces and it placed an unnecessary strain on British naval personnel. A more aggressive system of baiting its enemies into open-water battles would have allowed the Royal Navy to exercise its advantages and could have eliminated the threat that the French and Spanish fleets posed to Great Britain.

References

- ¹Maine, R. (1957) *Trafalgar: Napoleon's Naval Waterloo*. C. Scribner's Sons. Pg 254-258.
- ²Barrow, J. (1838) *The life of Richard, earl Howe, K. -G., admiral of the fleet, and general of marines*. J. Murray. Pg 216, 218-219.
- ³Bennet, G. (1977) *The Battle of Trafalgar*. Batsford. Pg 30.
- ⁴*Ibid.* Pg 23, 26.
- ⁵The term 'man-of-war' was used to describe all ships larger than frigates during the age of sail because of the line ahead formation that they traditionally employed in battle. 'Sail-of-the-line' and 'line-of-battle ships' were two other monikers that were used interchangeably to describe these vessels.
- ⁶Mahan, A. (1898) *The influence of seapower upon the French revolution and empire, 1793-1812*. Little, Brown & Co. Pg 194.
- ⁷Bennet, G. Pg 48.
- ⁸*Ibid.* Pg 48.
- ⁹Osborne, E. (2004) *Britain's economic blockade of Germany, 1914-1919*. Taylor & Francis. Pg 83-84
- ¹⁰Barrow, J. Pg 217-218.
- ¹¹*Ibid.* Pg 218.
- ¹²*Ibid.* Pg 233-234.
- ¹³Port St. Helens is located in the south of England.
- ¹⁴Barrow, J. Pg 230-232.
- ¹⁵To "strike a flag" means to surrender in naval terms. The phrase is taken literally from the act of removing the national flag from its place atop a mast.
- ¹⁶The term 'van' refers to the leading ships in a fleet.
- To 'haul wind' in nautical terms means to turn the front of the ship towards the point that the wind is blowing from. The French fleet performed this maneuver in order to expose their broadsides to the comparably-vulnerable fronts of the British ships.
- ¹⁷Barrow, J. Pg 230-232.
- ¹⁸*Ibid.* Pg 233-234.
- ¹⁹Collingwood, C., Collingwood, G. (1829) *A selection from the public and private correspondence of Vice-Admiral Lord Collingwood: interspersed with memoirs of his life*. G. & C. & H. Carvill. Pg 25;
- The names of ships often remained unchanged after they were captured, which is why many of the Royal Navy's vessels from the period in question had French or Spanish names.
- ²⁰Collingwood, C. Pg 3.
- ²¹To J. E. Blackett, Esq. *Excellent*, off Toulon, May 11, 1796 in *Ibid.* Pg 38.
- ²²To J. E. Blackett, Esq. *Excellent*, Still off Toulon, September 25 1796 in *Ibid.* Pg 42-44;
- To J. E. Blackett, Esq. *Excellent*, Gibraltar, December 5, 1796 in *Ibid.* Pg 44-45.
- ²³Carthage can also be spelled Cartagena. Both refer to the same city in Spain.
- ²⁴To J. E. Blackett, Esq. *Excellent*, off Cadiz, January 26, 1798 in Collingwood, C. Pg 84-87.
- ²⁵To J. E. Blackett, Esq. *Excellent*, off Cadiz, January 26, 1798 in *Ibid.* Pg 85-88.

- ²⁶To J. E. Blackett, Esq. *Triumph*, off Mahon, July 11, 1799 in *Ibid.* Pg. 104.
- ²⁷*Triumph*, Torbay, August 18, 1799 in *Ibid.* Pg. 105.
- ²⁸White, J. (1806) *Memoirs of the Professional Life of the late Most Noble Lord Horatio Nelson*. J. Cundee. Pg 106-120, 174-176.
- ²⁹*Ibid.* Pg 78.
- ³⁰Clarke, J., M'Arthur, J. (1810) *The life of Admiral Lord Nelson, K. B. from his lordships manuscripts*. Caldwell and Davies. Pg 451-452.
- ³¹*Ibid.* Pg 456.
- ³²*Ibid.* Pg 457.
- ³³White, J. (1806) *Memoirs of the Professional Life of the late Most Noble Lord Horatio Nelson*. J. Cundee. Pg 78-82.
- ³⁴*Ibid.* Pg 83-85.
- ³⁵*Ibid.* Pg 88-92.
- ³⁶*Ibid.* Pg 93-94.
- ³⁷*Ibid.* Pg 100-104.
- ³⁸*Ibid.* Pg 105.
- ³⁹*Ibid.* Pg 106-120.
- ⁴⁰*Ibid.* Pg 128.
- ⁴¹*Ibid.* Pg 96.
- ⁴²*Ibid.* Pg 173.
- ⁴³*Ibid.* Pg 169.
- ⁴⁴*Ibid.* Pg 174.
- ⁴⁵Clarke, J. Pg 606.
- ⁴⁶Bennet, G. Pg 138.
- 'The Nelson touch' initially referred to Lord Nelson's tactics for the Battle of Trafalgar, but later more generally referred to his particular leadership style.
- ⁴⁷Nelson had moved from his own ship, the *St. George*, to the smaller *Elephant* for the battle so that he could maneuver in the narrow Danish waters.
- ⁴⁸Clarke, J. Pg 606-608.
- ⁴⁹White, J. Pg 174-176.
- ⁵⁰Captain Puget to Cornwallis, *Foudroyant* Cawsand Bay, June 23, 1804 in Leyland, J. (1902) *Dispatches and letters relating to the blockade of Brest, 1803-1805*. Navy Records Society. Pg 5-6;
- Ibid.* Pg 3-4.
- ⁵¹*Ibid.* Pg 4.
- ⁵²*Ibid.* Pg 5-6
- ⁵³*Ibid.* Pg 2-8;
- Lord Melville to Cornwallis, Admiralty, July 10, 1804 in *Ibid.* Pg 8-9;
- Lord Melville to Cornwallis, Wimbledon, August 9, 1804 in *Ibid.* Pg 9-10;
- Memorandum by Captain Brisbane, *Glory*, August 16 1804, in *Ibid.* Pg 10-11;
- Cornwallis to Lord Melville, *Glory*, off St. Helen's, August 17, 1804 in *Ibid.* Pg 11-12.
- ⁵⁴*Ibid.* Pg 12;
- Lord Melville to Cornwallis, Wimbledon, August 26, 1804 in *Ibid.* Pg 13-14;
- Lord Melville to Cornwallis, Admiralty, September 5, 1804 in *Ibid.* Pg 14-15;
- Admiral Young to Cornwallis, Plymouth Dock, September 7, 1804 in *Ibid.* Pg 21.
- ⁵⁵Cornwallis to the lieutenants of the *Rifleman*, *Pelter*, *Biter*, *Thrasher*, *Phosphorus*, By the Hon. William Cornwallis, Admiral of the White &c. in *Ibid.* Pg 26;
- Cornwallis to Lieutenant Higginson (*Happy Return*) in *Ibid.* Pg 27.
- ⁵⁶Cornwallis to Lord Melville, 16 October 1804 in *Ibid.* Pg 29.
- ⁵⁷Barrow, J. Pg 218-219.
- ⁵⁸Brenton, E. (1838) *Life and correspondence of John, earl of St. Vincent*. H. Colburn. Pg 206-207.
- ⁵⁹An Officer of His Majesty's Navy (1804) *An answer to Mr. Pitt's attack upon Earl St. Vincent, and the Admiralty, in his motion for an enquiry into the state of the naval defence of the Country, on 15th of March, 1804*. H. Ebers. Pg 2.
- ⁶⁰*Ibid.* Pg 3-6.
- ⁶¹Mahan, A. Pg 111-115.
- ⁶²An Officer of His Majesty's Navy. Pg 6.
- ⁶³*Ibid.* Pg 6.
- ⁶⁴Colville, Q., Davey, J. (2013) *Nelson, Navy & Nation: the Royal Navy & the British people, 1688-1815*. Conway. Pg 148

- ⁶⁵Bennett, G. Pg 238.
- ⁶⁶Colville, Q. Pg 148, 208, 210, 219.
- ⁶⁷To W. Marsden, Esq. *Euryalus*, off Cape Trafalgar, October 22, 1805 in Collingwood, C. Pg 161-162.
- ⁶⁸*Ibid.* Pg 171.
- ⁶⁹*Ibid.* Pg 172.
- ⁷⁰*Ibid.* Pg 175.
- ⁷¹Bennet, G. Pg 191.
- ⁷²Beatty, W. (1825) *Authentic narrative of the death of Lord Nelson: comprising several interesting circumstances preceding, attending and subsequent to, that event*. W. Mason. Pg 25.
- ⁷³*Ibid.* Pg 28.
- ⁷⁴*Ibid.* Pg 29, 32.
- ⁷⁵To W. Marsden, Esq. *Euryalus*, off Cape Trafalgar, October 22, 1805 in Collingwood, C. Pg 163-164.
- ⁷⁶To J. E. Blackett, Esq. *Queen*, November 2, 1805 in *Ibid.* Pg 183-186.
- ⁷⁷Rodstock, W. (1806) *The British flag triumphant! or, the wooden walls of old England: being copies of the London gazettes, containing the accounts of the great victories and gallant exploits of the British fleets, during the last and present war; together with correct lists of the admirals and captains in the several engagements: to which is prefixed, an address to the officers, seamen, and marines, of His Majesty's fleets*. Rivington. Pg 3-4.
- ⁷⁸See pages 2 and 3.
- ⁷⁹To Mrs. Collingwood, *Dreadnought*, off Cadiz, August 21, 1805 in Collingwood, C. Pg 148-149.
- ⁸⁰From Lord Nelson, *Victory*, September 25, 1805 in *Ibid.* Pg 153.
- ⁸¹From Lord Nelson, *Victory*, October 10, 1805 in *Ibid.* Pg 157.
- ⁸²To Lord Nelson, October 6, 1805 in *Ibid.* Pg 155.
- ⁸³Mahan, A. Pg 184-185, 197.
- ⁸⁴*Ibid.* Pg 272, 290.
- To be denationalized was the naval equivalent of being declared an outlaw.
- ⁸⁵Mahan, A. Pg 272, 283-288.
- ⁸⁶Maine, R. Pg 254-258

A Song of Burning Wood: Colonization and Memory in *A Song of Ice and Fire*

BY HANNAH GILDART
SOUTHWESTERN UNIVERSITY

“Foreigners will rebuild your walls, and their kings will serve you.”
- Isaiah 60:10

For Jason Magnon-
You would have understood that
“History... is a nightmare from
which I am trying to awake.”
— James Joyce, *Ulysses*

Introduction

Background of colonization in A Song of Ice and Fire

George R.R. Martin's series *A Song of Ice and Fire* has outstripped many of its contemporaries in the fantasy genre in terms of popularity, in part because of the phenomenal success of its television adaptation, *Game of Thrones*. Viewers are captivated by the narratives of the country Westeros—also named the Seven Kingdoms—a fictional nation that parallels Medieval Europe. Part of the appeal that Martin provides is a world with a low moral standard. Readers and viewers explicitly relish the opportunism, selfishness, and violence of Westeros. Rather than focusing on the common folk or those displaced by colonial rule, George R.R. Martin focuses on aristocrats vying for power. However, there is an earlier wave of violence that has become little more than a backdrop for that action, but that provides context and resonance for it. The narratives of *A Song of Ice and Fire* are interwoven with Westerosi myths that date back to 12,000 years before the setting of the series, when the conquest of the Western continent by humans began. The colonization included attempted genocide and the displacement of indigenous non-human species, such as giants, “children of the forest”, direwolves, and the Others. The only narrative points of view accessible to the reader are those of humans; thus the reader/viewer is not expected to consider the points of view of indigenous groups. These myths are worthy of critical analysis not only because they are repressed histories that inform the text's present-day conflicts. More importantly, the myths offer readers the opportunity to identify and examine the ways in which *A Song of Ice and Fire* offers a paradigm of a typical process for colonization: displacement of indigenous populations, forgetting, and mythologizing (as a technique of forgetting).

This paper begins by describing the

displacement of indigenous peoples via a series of human invasions and informing the reader of the ways in which the efficacy of forgetting through mythologizing is configured in the series. The paper then frames the region North of Westeros as a foil to Westeros, and also explains the ways in which the Others are an unreal, yet, existential threat. “*A Song of Earth*” begins the recentring of the paper on the indigenous group, known to humans as “the children of the forest.” The “children” are the primary reason that the Northern region of Westeros has a different identity from the rest of the nation. Although Northerners are descendants of the human settlers who did not attempt genocide on the “children,” the North misinterpreted their culture. The discussion of the North's misinterpretation leads into an illustration of the extent to which the cultural/sacred work of “the children of the forest” is memory.

“Foreigners will rebuild your walls, and their kings will serve you”

One of the myths which is formative for Westeros is the myth/history of the Long Night, which happened approximately 12,000 years before the present narratives. This period is roughly double the human history the real world has, but *A Song of Ice and Fire* intentionally has a much deeper history than our world. This might be Martin's response to Modernity's ephemeral perception of history: his individual backlash against “our hopelessly forgetful modern societies” which “organize the past” through the dilated lens of history.¹ Theorist Pierre Nora explains what Modernity has done to historical perception:

Indeed, we have seen the tremendous dilation of our very mode of historical perception, which, with the help of the media, has substituted for a memory entwined in the intimacy of a collective heritage the ephemeral film of current events.

The “acceleration of history,” then, confronts us with the brutal realization of the difference between real memory-social and unviolated, exemplified in but also retained as the secret of so-called primitive or archaic societies-and history, which is how our hopelessly forgetful modern societies, propelled by change, organize the past.¹ In *A Song of Ice and Fire*, Martin insists upon

a world that does not forget as quickly as our modernity. While Westeros does repress memories of millennia past into myths, these myths survive, and provide a window into the collective unconscious of certain groups in the present. The Northern region of Westeros remembers the most about the Long Night, and displaced groups such as “children of the forest”, the Others, direwolves, and the Giants.

The two characters from which we hear most about myths of the “children of the forest” are Old Nan and Maester Luwin, who live in the North of Westeros. Old Nan is the wet nurse that cared for three generations of Starks. It is noteworthy that we know her tales from the perspective of Bran Stark, who will join the “children of the forest.” Old Nan tells Bran the story of the Last Hero, one of the First Men who lived during the Long Night, a Winter which lasted a generation. The myth of the Long Night describes the designation of certain groups as “other” (that which is abject: separated from the self in order to maintain a coherent subject), and define the boundaries between normativity (what me might call “Westerotivity”) and otherness.²

The Long Night took place in the Age of Heroes, the era which began after the First Men settled on the Western land mass, once the First Men and the “children” brokered a peace. The Long Night is crucially characterized by the swarming of the Western continent by the Others, who were killing humans *en masse*. The Others, indigenous undead and undying beings, essentially waged an apocalypse against humanity. Profoundly threatened, the humans had the Wall built to ostensibly keep the Others confined to the desolate Northernmost region of the continent. The myths suggest that Brandon the Builder, founder of house Stark, enlisted the giants and the “children of the forest” to help build the Wall. The “children” would have imbued the Wall with their magic to make it impenetrable to the Others. Regardless of the “children's” and giant's generosity, the humans' reactionary panic caused them to feel abjectly threatened by all groups outside of Westerosi boundaries of normalcy, not just by the Others. Thus, in the process of displacing the Others, the humans displaced all other non-normative groups as well, such as the “children,” giants, and some humans who

were unable to conform to Westerosi normalcy.

Old Nan's tales: abject others abjured

Much emphasis is put on Old Nan's age, and readers are encouraged to see her as having more connections to the memory of the ancient past in which non-human species coexisted with the First Men. The problem is, although Old Nan gives no indication that these narratives are false, no one expects Old Nan's narratives to be more than myth. The fact that the narratives she provides are called "stories" and are not granted the same level of credibility as Maester Luwin's narratives is one of many examples of the severely patriarchal culture with which Martin structures the world of *A Song of Ice and Fire*. Old Nan's tales are misread and their subversive force ignored. That Westeros sees narratives such as Old Nan's as myth allows present Westeros to unconsciously continue the erasure of the ancient past. After being separated by the Wall for approximately 8,000 years, many humans believe that the "children of the forest" were never more than myth. This erases the possibility of acknowledging their influences on human cultures as well as almost all of the "children's" actual existence and prevents any reparations to the "children of the forest." This is exactly how conquest functions in the real world: colonization depends on the process of displacement and forgetting. In discussing Ursula LeGuin's fantasy/sci-fi novel *City of Illusions*, notable author Judith Caesar offers an observation that I argue starkly echoes the process of forgetting which happened to Westeros over time:

In *City of Illusions*, mind-razing, the destruction of a person's memories is developed as a metaphor for the process of colonizing the minds of a subject people. It suggests the ways in which history can be ignored, destroyed, or misinterpreted; the ways in which cultural values can be distorted and misrepresented.³

Caesar recognizes that *City of Illusions* illustrates the ways in which understanding and manipulating the thoughts of the members of one's own group as well as the colonized group is a powerful tactic of colonization.

In *A Song of Fire and Ice*, narratives told by Westerosi people to other Westerosi people destroy knowledge of the colonized societies. This manipulation of knowledge functions the way Judith Caesar argues mind-razing functions: the Westerosi raze human minds of the "children." Creating master narratives that are hostile to indigenous peoples or erase them outright is a characteristic of colonization. The colonizers use a double process of

silencing and forgetting. Myth becomes an opportunity to forget: to naturalize an order. Memory, for theorist Pierre Nora, is authoritative for each group that depends on it.¹ History, for Nora, is in conflict with memory. History, because it pretends to be objective, destroys memories of colonized groups. History and archiving are techniques developed in response to the disappearance of memory, and are in fact, techniques of forgetting. Indeed, mythologizing is a form of remembering that enables the hearers to forget. Both *City of Illusions* and *A Song of Ice and Fire* show us that mythologizing is an effective tool for ensuring that the colonized societies stay displaced. Yet, the texts create opportunities for remembering, creating a parallel and counterweight for the vehemence and the violence of the present.

These myths work to construct Westerosi identity and, in some cases, particular regional Westerosi groups. The apocalypse and subsequent building of the Wall left a legacy in contemporary Westeros: The Wall is used as a symbol of national identity by constructing the North of the Wall as a foil to Westeros. The Northern Westeros-Northernmost wilderness foil serves as a mirroring structure that recalls Joseph Conrad's discourse of Africa in his novel *Heart of Darkness*, another work of literature which concerns itself with colonialism. Renowned author Chinua Achebe sees *Heart of Darkness* as participating in a larger Western cultural discourse on Africa. "Quite simply it is the desire—one might say the need—in Western psychology to set Africa up as a foil to Europe, a place of negations at once remote and vaguely familiar in comparison with which Europe's own state of spiritual grace will be manifest."⁴ In like fashion, "North of the Wall" is what Westeros is not. The collective lives of the people of Westeros constitutes normalcy and Civilization; North of the Wall constitutes the Other.

For present day Westeros, the Wall evokes memories of before the Wall was built, when abject otherness rampaged across the Western continent. Thus, the Wall is a symbol of national identity because it is a visual manifestation and reminder of the illusory security and normalcy Westeros felt it needed after it was built. The hostile Northern Wilderness is spoken of with leftover horror of the apocalypse because it represents another potential apocalypse to Westeros. Because the North of the Wall is at once remote and vaguely familiar, Westerosi make it a space of negation. "In the Seven Kingdoms it was said that the Wall marked the end of the world" so that the possibilities of the reality there can be forgotten.⁵

As their name suggests, the Others

represent the epitome of this foil to Western identity. They represent a particular kind of otherness, one that is nightmarishly apocalyptic in nature. They are the epitome because not only are they dangerous to humans, but their otherness was so enormous in scope that it almost, and still has the potential to, exterminate humanity. Because Westerosi people have not seen an Other since the Wall was built, Westeros is able to repress memories of the Others and the near-genocide they performed. Despite the Wall being 700 feet of evidence that something horrible occurred to cause its creation, the Others were mythologized to preserve Westeros' state of normalcy. If Westeros had continued to acknowledge the Others' existence, they would have had to accept the idea of another potential end of humanity. In Freudian terminology, the Others are *unheimlich*, a word which describes something that is horrifying by virtue of being familiar: The horror comes from the threat of the return of repressed memories.⁶ The idea of the Others brings repressed memories of human extinction to the surface, which Westeros can only handle as myth, lest its perception of the longevity of humanity fall apart.

Yet, on some unconscious level, Westerosi are aware that the Others exist, even if the horror of their existence is largely ameliorated by mythologizing the Others. The Others are even more *unheimlich* because part of what has been repressed about the Others is that they were humans who were transformed into the Others by the "children of the forest" to protect themselves from the humans. Julia Kristeva, who has done much work on the ways in which our culture constructs the self and other, explains abjection:

The abject has only one quality of the object—that of being opposed to I. If the object, however, through its opposition, settles me within the fragile texture of a desire for meaning, which, as a matter of fact, makes me ceaselessly and infinitely homologous to it, what is *abject*, on the contrary, the jettisoned object, is radically excluded and draws me toward the place where meaning collapses.⁷

Kristeva's definition of the abject is applicable to the construction of the Others because the abject refers to that which threatens collapse of meaning caused by loss of distinction between the self and other.⁷ Westerosi humans do not even have to know about the common genetic origin of the Others for the Others to be abject. The Others are already similar to humans because both humans and the Others are genocidal and apocalyptic, depending on perspective. Trying to protect themselves from dissolution, the *super-egos* of

Westerosi reject memories of the common (in)humanity of the Others, who would remind humans of their own genocidal *Id*.⁶

A Song of Earth

The readers of *A Song of Ice and Fire* do not hear the perspective of the otherized and displaced groups. Rather, the reader knows of these groups from Westerosi myths, thus the reader is not expected to consider the points of view of indigenous otherized groups. However, this paper works to denaturalize the narratives of the humans of Westeros, in part to suggest that viewing and reading *A Song of Ice and Fire* from the perspective of displaced groups—particularly the perspective of the “children of the forest”—is a valuable practice.

In the most recent book in the series, *A Dance of Dragons*, we meet the “children” through Bran Stark, who travels north of the Wall to meet them. The “children” tell Bran that, in forming a group identity, the name they choose for themselves is “*Those Who Sing the Song of Earth*,” and that they also accept the name “*Singers*.”⁸ Until this point, we have known them as “children” because when humans began settling on the Western continent they imposed that name. Although “*Those Who Sing the Song of Earth*” is only a translation of the name they have for themselves into the Common Tongue, using the names they use for themselves is a necessary first step in centering the margin because the human’s renaming erased their ability to represent themselves. The renaming is also anthropocentric, or “human-centered,” in the sense that it defines *Those Who Sing the Song of Earth* by their physical difference from humans. *The Singers* are about the size of a human child, thus by comparing them to human children through naming, the term reveals itself as human-centered. By changing the name of *Those Who Sing the Song of Earth* to “children,” (e.g., beings not capable of autonomy) humans manipulated the knowledge of *the Singer’s* group identity and of their individual identities as adults with agency. Although it was clear to humans which *Singers* were adults, they knowingly referred to adults as “children,” which denies the personhood and agency of adult *Singers*. Therefore, humans manipulated knowledge by constructing *Those Who Sing the Song of Earth* as children: inferior to adult humans within the value system of humanity.

The title *A Song of Ice and Fire* refers to the prioritized Westero-centric narratives of aristocrats vying for control. In contrast, songs of Earth consist of narratives of being indigenous or of being one of *Those Who Sing the Song of Earth*, which/who are necessarily obscured by virtue of the title of the series being *A Song of Ice and Fire*. Following the

pattern of title as a metaphor for narrative, this paper is *A Song of Burning Wood* because it is largely about the colonization of *Those Who Sing the Song of Earth* and because it provides a different way to read the title and meaning of the series itself.

Before the *Singers* were displaced, their identity as *Those Who Sing the Song of Earth* (in contrast to “the children of the forest”) was self-evident. The second wave of invaders, the Andals, forced the exposure of *Those Who Sing the Song of Earth* into contact with human processes of history. *Those Who Sing the Song of Earth* have always been “peoples of memory,” but before the Andals, they were able to live within memory.¹ Since the destruction of the self-evident nature of their memory (and most of their actual memory), the preservation of their remaining memory has become emphatically important. The invasion of the Andals represents:

A particular historical moment, a turning point where consciousness of a break with the past is bound up with the sense that memory has been torn-but torn in such a way as to pose the problem of the embodiment of memory in certain sites where a sense of historical continuity persists. There are *lieux de memoire*, sites of memory, because there are no longer *milieux de memoire*, real environments of memory.¹ The remaining communities of *Those Who Sing the Song of Earth* and the memorials of their culture constitute sites of memory because they no longer have access to their real environments of memory: the entire Western continent before the conquest of humans.

Deforest and conquer: “The North Remembers” imperfectly at best

Readers are often reminded that the blood of the First Men runs through the Stark veins, and much of the population of Northerners. This means that they are descendants of The First Men, the first group of colonizers, who did not engage in genocide, but were still invaders. They invaded and settled on the Western land mass, but they did not kill many of the *Singers*, and, just as crucially, they did not burn down the weirwood trees (trees which formed the *Singers’* culture). And although they were more successful at gaining territory than *Those Who Sing the Song of Earth* were at keeping theirs, this process took thousands of years, during which time The First Men were influenced by *Those Who Sing the Song of Earth* and brokered a peace with them. The North was the penultimate kingdom on the Western continent to lose their independence and become part of Westeros. The Northern phrase “the North Remembers” perhaps refers to the thousands of years of separation

from the other kingdoms of Westeros, during which time their own regional identity flourished. It could also refer to the time before the Andals murdered most of *Those Who Sing the Song of Earth*, from whom the North developed certain pivotal points of their identity, such as their religion and narratives. The North is constructed as a place that is not as disenchanting as the rest of Westeros. This state of enchantment causes the other six kingdoms to see the North as shrouded in ominous and backwards traditions. The cause of Westerosi prejudice against Northerners is that most Westerosi believe that the North excessively preserves old narratives and traditions. Westeros is prejudiced against Northerners because, to the rest of Westeros, the existence of The Northerner’s different perspective and identity undermines Westerosi perception of their culture as natural.

Maester Luwin tells Bran Stark that the First Men adopted the faith in the Old Gods from *Those Who Sing the Song of Earth*. Even when the Andals conquered most of the First Men’s land, the First Men prayed to the gods of the *Singers*, and that is why house Stark and other houses in the North keep the *Singers’* gods, now called the Old Gods, in the present. This is the major difference between Northern and non-Northern Westeros, although religion is very valuable to most characters in the series. Yet, it is acknowledged that faith in the Old Gods runs deeper, and is practiced with more gravity than the non-Northerners worship The Seven (New Gods). In order to pray to the Old Gods, The Northerners preserved some weirwood trees, which are fundamental to the culture of *Those Who Sing the Song of Earth*.

One of the *Singers* (the one Bran nicknames “Leaf”) explains to Bran that her people and the weirwood trees used to thrive all over what became Westeros. When the second wave of colonizers, the Andals, “burnt down the weirwood groves, [and] hacked down the faces”, they mostly succeeded in destroying *Those Who Sing the Song of Earth’s* collective knowledge, memory, and method of empathizing with other communities of *Those Who Sing the Song of Earth*.⁹ The burning of the weirwood trees is akin to the actual colonial tool of burning and falsifying the documents of indigenous people because both exemplify the “rewriting or suppressing their history.”³

Even though the Old Gods are worshipped with sincerity, and although the First Men did not kill many of *Those Who Sing the Song of Earth*, the adoption of certain parts of the *Singers’* culture constitutes another manipulation of knowledge by humans. Although in this instance the manipulation of knowledge was

accidental, and the Northerners have always had a genuinely profound reverence for the Old Gods. Still, the North only partially remembers the religion of *Those Who Sing the Song of Earth* because by virtue of not being *Singers*, the Northerners could never have learned the full function of the weirwood trees or the full extent of their relationship to *Singers* and the Old Gods. The North's use of parts of the *Singers'* culture suggests the "way cultural values can be misrepresented," in that Northern humans use weirwood trees as altars to pray to the Old Gods, and believe that they are sacred to the Old Gods.³ To *Those Who Sing*, the weirwood trees are the Old Gods. Northerners did not misrepresent the *Singers'* culture with the intention of undermining it. They misrepresented it because the full nature of the weirwoods and *Those Who Sing* is unintelligible to almost all humans. Not only are the trees literally the Old Gods, but they are a telepathic way for *Those Who Sing* to preserve their memory and culture. This misinterpretation of the weirwoods is echoed by Northerners giving the weirwood trees a much less *unheimlich* name: the Heart Trees.

The Singers remembered

The weirwood trees function as conduits for the consciousness of shamans called greenseers through which they could access the memory, and experience the present from the perspective of the weirwood trees. Thus they could communicate with other communities of *Those Who Sing the Song of Earth*, and access memories of the greenseers. In this way, the weirwood trees are like the literature of *Those Who Sing the Song of Earth*, and thus Le Guin's mind-speech, a form of empathy and telepathy, which "gives the power to understand the thoughts and feelings of others, as a common literary language does when combined with literary sensitivity and empathy."³ Mind-speech functions in *City of Illusions* as literary language functions in the real world. At the very least, greenseeing functions in the way Caesar argues mind-speech functions. In *A Song of Ice and Fire*, the character Jojen from compares greenseeing to literature:

A reader lives a thousand lives before he dies. The man who never reads lives only one. The singers of the forest had no books. No ink, no parchment, no written language. Instead they had the trees, and the weirwoods above all. When they died, they went into the wood, into leaf and limb and root, and the trees remembered. All their songs and spells, their histories and prayers, everything they knew about this world. Masters will tell you that the weirwoods

are sacred to the old gods. The singers believe they are the old gods. When singers die they become part of the godhood.⁸

Although Jojen only compares greenseeing to literature, literature is only a part of the interconnected web of meaning that makes up culture. The *Singers* did not need written language, or any means of archival because their lives were memory: There was no distinction between the past and the present. Rather, the present was a constant living of memory, and this was the essence of their entire culture. Pierre Nora would characterize the *Singers'* culture as Pre-Modern, because the acceleration of history had not yet swept away their memory.¹ The culture of *Those Who Sing the Song of Earth* was born of memory; their culture was founded in its name.¹ However, as the number of *Those Who Sing the Song of Earth* and their weirwood groves dwindled, they began to fall prey to the stagnation and isolation that LeGuin represents in *City of Illusions*. It is not clear whether the weirwood trees have always appeared the way they do in the present, with faces, white bark, red leaves, and red sap coming from the eyes which resembles blood. Perhaps the white bark represents the memories of the shock of attempted genocide: as if the weirwood beings blanched with the murdered *Singers*. The fact that they appear to be crying blood is an embodiment of the grief which the trees help process as memorials.

Winter is coming

Interestingly, Martin's trademark irony shows how the colonizers tools can backfire. Westeros's practice of erasure has been in the process of backfiring for 8,000 years, and begins to become visible to certain humans in the present setting. The human's use of erasure works in two ways. The physical genocide of *Those Who Sing the Song of Earth* caused them to protect themselves: They weaponized human beings by transforming some of them into the Others. Westeros's propagation of the belief that non-human peoples never existed allowed the Other's army to build itself up to an indomitable size. The Brothers of Night's Watch repeatedly ask every house for more men to join them at the Wall because of signs that the numbers of the Others are growing. They invariably refuse because the Others, like all non-human peoples, were never more than myth. Indeed, the humans' technique of colonization enables their destruction.

Even the existence of racial "others" in Westeros threatens the nation's boundaries of the self. Constituting certain peoples as "alien" is the most effective way of forgetting that humans invaded Westeros. "This appears

as an expulsion of alien elements, but the alien is effectively established through this expulsion".² Expulsions through the building of walls is the most effective way for dominant groups to forget the common humanity between themselves, and the groups they have displaced.

Perhaps thinking about the colonial past, present, and future is terrifying, and processing the horror we feel is more possible through fantasy, whether the process of trying to understand colonialism is conscious or unconscious.

References

- ¹Nora, P. (1989) "Between Memory and History: *Les Lieux de Mémoire*." *Representations* 26. Pg 7-24.
- ²Butler, J. (1999) *Gender Trouble: Feminism and the Subversion of Identity*. Routledge Classics.
- ³Caesar, J. (2011) "City of Illusions, Layers of deceit, and the Sterility of Isolation." *Contemporary Justice Review* 14.2. Pg 227-246.
- ⁴Achebe, C. (2012) "An Image of Africa: and, the Trouble with Nigeria." Penguin Random House.
- ⁵Martin, G.R.R. (2000) *A Storm of Swords*. Bantam Spectra.
- ⁶Freud, S. (1956-1974) *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. Hogarth Press.
- ⁷Julia, K. and Roudiez, L. (1982) *Powers of Horror: An Essay on Abjection*. Columbia University Press.
- ⁸Martin, G.R.R. (2011) *A Dance of Dragons*. New York: Bantam Spectra.
- ⁹Martin, G.R.R. (1996) *A Game of Thrones*. New York: Bantam Spectra.

And There They Were: Mrs. Dalloway, British Identity and the Indian Other

BY JOELLE HAMILTON
COLORADO STATE UNIVERSITY

In Virginia Woolf's novel *Mrs. Dalloway*, the sign 'India' floats on administrative gossip, alongside Miss Kilman's traveling trunks, and in the memory of Helena Parry. The reader must look at the fringes of the parties, the sugared almonds, and the green silk dresses, and ask what a novel ostensibly about a single day in London might have to do with India, itself more a political concept than an ethnic reality. What could London have to do with the geographically and culturally distant subcontinent of India? Furthermore, why might it be important that in *Mrs. Dalloway*, India is never fully present, but rather endures as a memory in the minds of the characters who have both recently returned from India and remember India from before the century's end? Using postcolonial critic Edward Said's concept of orientalism, which is, broadly, Western discourses or "system[s] of knowledge about the East," I hope to examine how the novel's characters construct British identity through and in opposition to the Indian other.¹ When India does appear in the novel, it is a flattened, abstract concept associated both with ignorance and the exotic. By understanding India as a term that involves a complex set of discourses that draw an "ontological and epistemological distinction" between the East and the West, we can better understand India as reflected in the occidental, imperialist perspectives of the novel's characters.¹ However, just because *Mrs. Dalloway* affects and is also affected by problematic colonial representations of India does not mean that it is not valuable both as a historical artifact and a work of literature. I will argue that through *Mrs. Dalloway*'s free indirect discourse, which focuses closely on individual experience and identity construction, the novel reminds us what it means to reconstruct the past and what it means to remember and be remembered.

Through its narrative technique, *Mrs. Dalloway* seems to work towards what Indian historian Gajendra Singh advocates for in writing "life-history": a focus on the individual or witness that "takes into account failings of memory and everyday myth-making."² The life-histories of Peter Walsh, Helena Parry, and Clarissa Dalloway question what it means to be remembered, as they construct and are constructed through each other's faculties of memory and acts of speech. The novel's narrative technique brings to light what Singh in his book *The Testimonies of Indian Soldiers and the Two World Wars* argues is essential to writing life-history,

and especially applies to the context of subaltern studies of South Asia: thinking about the "experiences and requirement of the individual—how the person copes with society rather than how society copes with the stream of individuals."² *Mrs. Dalloway*, in its close-up focus on the lived experiences of its characters, embodies such a narrative history.

In writing both history and literature, some voices will remain silent and forgotten. Positing a complete history presumes an inevitable and knowable past, and thus ignores the essential fact of contingency, loss, and absence inherent in writing narrative histories. I will start at the end of things and employ a non-chronological approach some historians use to highlight the contingency of history.² At the end of *Mrs. Dalloway*, Clarissa Dalloway, hosting her much-anticipated party, leads Peter Walsh to her Aunt Helena Parry to converse over their shared experiences in India:

At the mention of India, or even Ceylon, [Miss Parry's] eyes (only one was glass) slowly deepened, became blue, beheld, not human beings — she had no tender memories, no proud illusions about Viceroy, Generals, Mutinies — it was orchids she saw, and mountain passes and herself carried on the backs of coolies in the 'sixties over solitary peaks; or descending to uproot orchids (startling blossoms, never beheld before) which she painted in water-colour; an indomitable Englishwoman, fretful if disturbed by the War, say, which dropped a bomb at her very door, from her deep meditation over orchids and her own figure journeying in the 'sixties in India — but here was Peter.³

Helena Parry reproduces a European Imperialist myth of empty, unpossessed land as she imagines the Burmese landscape (today Myanmar) as "solitary" and seemingly populated only by exotic orchids. The "coolies" (a derogatory term for Asian laborers) she refers to are literally *beneath* her and do not seem to count as people, as she places them outside the category of human beings—"her eyes [...] beheld, not human beings... it was orchids she saw."³ Helena uses the gaze of the colonizer to conceive of an India not as a space of culture, local history, and living and feeling human beings but rather as a novel canvas for the self-construction of her own knowledge and identity; it is through her exclusion of the Indian other

that Helena gains agency as botanist and an "indomitable Englishwoman."³ Gaze functions interestingly here, for Helena, with her one glass eye, constructs the indigenous as sightless—the orchids she paints are "never beheld before," implying that her sight counts as the first discovery of these flowers. Helena's watercolors form a scientific system of knowledge about India as an unpopulated, exotic locale, where indigenous bodies lack sight and agency, and function solely as a mode of travel.

It is interesting what Helena explicitly excludes from her memory of Burma: she does not see "human beings"—here the rough opposition of colonizing and colonized actors of Indian imperial politics—Viceroy, Generals, Mutinies," and thus she has "no tender memories" or "proud illusions."³ The proper nouns Helena lists follow one right after another, gesturing to a loosely conceived hierarchical structuring of British rule: colonial governors first, then generals, and lastly "Mutinies," which are not individual actors, but events, specifically armed uprisings. The people of India are not represented bodily in this schema; they are simply a collective situated at the bottom of a colonial hierarchy. In constructed opposition to the ruling hierarchy of Britain, through their acts of defiance, Indians are created as the other.

By focusing on the non-human elements of India—the exotic and natural beauty of the landscape—Helena's memory of India claims an essentially apolitical position unaffected by "illusion."³ While she claims a perspective outside of political discourse, outside of the administrative affairs of colonial India, with access to a "true" image of India; Helena nonetheless, through her coded position as superior to the Indian other, participates in the very orientalist discourses she eschews. In *Orientalism*, Said contends that the Occident's conception of the Orient is not simply fantasy, or a product wholly of the imagination. Said argues that the system of knowledge about the Orient is importantly founded in reality; however, the culture of the Orient remains a larger reality than orientalist discourses can contain or understand, for the Orient's "lives, histories, and customs have a brute reality obviously greater than anything that could be said about them by the West."¹ At the end of this passage, Helena Parry gains embodiment as a result of her memory of India; she sees her "own figure journeying in the 'sixties

in India.”²³ In this way, Helena constructs herself as an indomitable artist and an embodied individual through her memory of Burma in contrast to the unembodied, sightless inhabitants of India.

A similar action of British identity construction takes place early in the novel, when Peter Walsh, who has recently returned from India, recalls his time in India as an administrator whilst looking at his image in a shop window:

And there he was, this fortunate man, himself, reflected in the plate-glass window of a motor-car manufacturer in Victoria Street. All India lay behind him; plains, mountains; epidemics of cholera; a district twice as big as Ireland; decisions he had come to alone — he, Peter Walsh; who was now really for the first time in his life, in love. [...] For he had a turn for mechanics; had invented a plough in his district, had ordered wheel-barrow from England, but the coolies wouldn't use them, all of which Clarissa knew nothing whatever about.³

The free indirect discourse style here begins with a definitive statement of Peter's spatial and temporal existence that will be echoed throughout the novel—“And there he was.”²³ This statement of existence enacted through the passive past tense delineates Peter's individuality as he situates himself in both time and space. Peter foregrounds himself as a unique individual juxtaposed against a background of a removed Indian; he imagines India as physically behind him, both in space, as if he is eclipsing the entire continent, and also temporally, as India now persists in his memory. At first, Peter's description of India is one of vastness, the scope of limitless landscape conjured by unqualified “plains” and “mountains.” From this panorama we move to quite another kind of sweeping vastness: the viral spread of disease. Cholera, a disease beginning with fever-like symptoms and ultimately deteriorating the body to the point of death, heavily preoccupied colonial administrators; the origins and transmission of the disease were unknown and British administrators were tasked with preventing and containing outbreaks. Additionally, before 1868, cholera in British medical discourses was thought to be “endemic (constantly present) at the mouth of the River Ganges, in Bengal, in India.”²⁴ On this understanding, medical and administrative discourses about cholera associated India and the colonized body with the spread of deadly, infectious disease. The myth of cholera's single, geographically specific origin in India speaks to how the political realities of colonial rule can affect medical knowledge. Thus Peter, in

his depiction of a cholera-swept India, works within a long-standing colonial framework that understood India as the source of epidemic disease; Peter expresses anxiety about his identity as British colonial administrator and his fears about the spread of disease.

Peter also enacts a conflation of his administrative district with the whole country of India: “All of India lay behind him...a district twice as big as Ireland.”²³ The whole area of British-ruled India, which would in 1922 include modern-day countries Myanmar, Bangladesh, Sri Lanka, Pakistan, and Bhutan, is reduced to Peter's administrative district. In this way, India serves as a sign that confers on Peter his status as colonial administrator, and functions to confirm his own masculine, British image reflected back to him in the glass. He depicts the “coolies” in his district as obstinate and backwards, and Peter voices his frustration at their refusal to use his imported wheelbarrows and invented machine. Peter associates himself with an ingenuity and individuality that he believes the Indians do not possess; indeed, the “coolies” do not utilize Peter's contributions. Through his depiction of a cholera-swept India metonymically representing his own administrative district, Peter constructs himself as resourceful, solitary, and unique, and reproduces an orientalist dichotomy between Britain and India, West and East. He can be read constructing his identity through a relation to the colonized “coolie” Indian, the landscape of India, and the viral uncontrollable spread of disease. Like Helena, Peter places himself in “positional superiority” to the “coolies” he administers and thus reproduces orientalist discourses.¹

However, *Mrs. Dalloway*, through its narrative technique of free indirect discourse, also works at cross-purposes to its characters' constructions of identity through the Indian other. As the characters turn their lenses on each other, an interesting absence of identity takes place, particularly in the relationship between Peter and Clarissa. Clarissa remembers Peter, who has been absent for five years in the first page of the novel, as singularly unembodied and almost, in a way, in the midst of a disappearance. Looking out the window with a sense of foreboding, Clarissa imperfectly recalls Peter talking to her:

“Musing among the vegetables?”—was that it? — “I prefer men to cauliflowers”—was that it? He must have said it at breakfast one morning when she had gone out on to the terrace — Peter Walsh. He would be back from India one of these days, June or July, she forgot which, for his letters were awfully dull; it was his

sayings one remembered; his eyes, his pocket-knife, his smile, his grumpiness and, when millions of things had utterly vanished — how strange it was! — a few sayings like this about cabbages.³

In our first introduction to Peter Walsh, shakily filtered through Clarissa's unspoken reminiscence, Clarissa is unable to remember the Peter's exact words and instead offers us a series of possibilities: “Musing among the vegetables?”—was that it?—“I prefer men to cauliflowers?”—was that it?”³ Clarissa's memory, in her failure to land upon a singular and precise past, unseats and unfixes an objective, unified past. Clarissa, far from the steady and reliable omniscience offered through third person narration, cannot reconstruct another character. This opening scene sets up the imperfect, fractured logic of memory at play in the pages that will follow. Save for the occasional structuring interludes of third person narration that move the reader between individual perspectives, the free and indirect discourse places us within such fractured, faulty perspectives. Indeed, we do not receive a fixed description of Peter's physicality; rather, Clarissa's “remembered” Peter is a jumbled set of speech acts, physical features, personal effects, and personality traits: “it was his saying one remembered; his eyes, his pocket-knife, his smile, his grumpiness.”³ In this way, Peter, because of his proximal distance from Clarissa in both time and place seems to lose his identity through Clarissa's remembrance, the very identity he labors to construct throughout the novel.

Thus, while the characters in the novel seem to define, construct, and embody themselves in relation to an abstract, orientalist understanding of colonial India, another interesting action occurs in the novel as memory erodes identity, and uncertainty and possibility deconstruct surety of existence and presence. While the characters preoccupy themselves with constructions of identity through disembodied and othered representations of Indians, they also struggle with their own identities as incomplete and fractured. The characters' identity constructions both defiantly insist on fixed presence and enact a haunting disappearance. When Clarissa declares at her party both “there was old Miss Parry” and “here was Peter,” and when Peter speaks the final sentence of the novel “For there she was,” these declarations seem to be imbued with a sense of their fragility, self-aware of the anxiety underlying their speakers' utterances, haunted by the unspoken knowledge of imminent disappearance.³

Mrs. Dalloway simultaneously constructs identity through its characters' representations of an othered India and erodes these identities

through its characters' memories. The novel can be read as reinforcing orientalism in its creation of a divide between East and West, but can also be read as simultaneously deconstructing this binary. Through its free indirect discourse, the novel also reminds us about the nature of writing history, and speaks to a history of—to use Gajendra Singh's term—"haunting presences," where the individual's way of creating themselves and others is laced with anxiety about being remembered.² *Mrs. Dalloway* illustrates the desire of its characters to be wholly represented—to construct themselves as complete and whole in contrast to the other—and betrays an anxiety about disappearance underlying these desires.

Gajendra Singh, in his postcolonial history of sepoy participation in World War I, utilizes such an approach to reading the past. He reads for silences and for the latent desires of the sepoy soldiers and in doing so, asks for these forgotten witnesses of the Great War to be read and remembered.² While it is possible to argue that the past is unrepresentable because of our faulty faculties of memory, unvoiced desires, and large silences, Singh incorporates such ghostly accounts into his project of writing a sepoy life-history of World War I. He views history neither as objective and inevitable nor as disappeared and silent, but rather as a complex narrative emerging in the interplay of individual experience and complex political discourse. *Mrs. Dalloway* and Singh, are interestingly related in their representations of memory. Both could be read at odds with each other in their projects of representation, with *Mrs. Dalloway* enacting an effacement and Gajendra Singh seeking a recovery of Indian voices.²

In order to read the past and write history, as readers and historians, we must take into account our own faulty faculties of memory and unvoiced desires to be remembered as whole and complete. *Mrs. Dalloway*, a novel at once symptomatic of orientalist discourses and problematic in its understandings of the Indian other, is important to a project of writing life-history, for through the novel, we come to understand the anxiety of the individual's self-construction, and how the "other" functions to relieve this anxiety. Through Virginia Woolf's *Mrs. Dalloway* and Singh's project of sepoy life-histories, we come to understand the insecure individual as intimately concerned in any project of history. The individual is caught up in political discourses, at times unable to remember, desirous of fixed presence and identity, yet, all the while admitting absence. We learn to work with history not as complete and inevitable, but as incomplete and tenuous as our own individual memory.

References

- ¹Said, E. W. (2000) "Orientalism." In *The Edward Said Reader*. Rubin, A., eds. Vintage. Pg 69, 71, 73
- ²Singh, G. (2014) *The Testimonies of Indian Soldiers and the Two World Wars: Between Self and Sepoy*. Bloomsbury Academic. Pg 187, 190.
- ³Woolf, V. (1925) *Mrs. Dalloway*. Brace. Pg 3, 48-9, 178, 194.
- ⁴Watts, S. (2001) "From Rapid Change to Stasis: Official Responses to Cholera in British-ruled India and Egypt: 1860 to C. 1921". *Journal of World History* 12.2. Pg. 321.

Picking the Pretty One: A Qualitative Study of Toy Selection Among Girls Age Four to Ten

BY SAVANNAH ANDERSON-BLED SOE AND DONNA S. SHEPERIS, PHD.

LAMAR UNIVERSITY

ADRIANA DYURICH

M.S. Texas A&M University - Corpus Christi

Abstract

Children learn gender and social role types through a variety of influences including caregivers, media, and peers. Toy selection, particularly the choices of dolls by young girls, is also largely influenced by the child's concept of beauty.¹ Do they choose what looks most like them or what looks most like what they think is pretty? Are these the same or different concepts? Understanding the influences on preschool age girls is critical to educators working with this population. This research focused on the foundations of a qualitative study that looked into the following questions: What factors influence doll choice in girls age four to ten? If girls have access to a broad range of doll choices, will they tend to prefer a doll more similar to their ethnic or racial backgrounds or will they tend to choose the doll that is a more racially popular (i.e. socially acceptable) choice? It was hypothesized that young girls who have access to dolls that display the diverse range of skin colors, hair types, and facial features of various racial backgrounds, are less likely to choose a doll that is more socially acceptable (i.e. one that represents a white racial background). This qualitative content analysis study derived data from interviews with girls age four to ten in southeast Texas. The interviews were transcribed and analyzed for themes to determine the process used by girls in that sample to make choices about doll selection based on beauty. After the interview data was coded, several themes emerged. Overall, the factors that determined doll choice were: the doll's similarity in appearance to an external female role model, the doll's similarity in appearance to the participants, the appearance of the doll's hair, skin, and eyes, the doll's posture, whether the doll looked realistic, and whether the doll was smiling. Non-realistic dolls were deemed ugly, and dolls that were smiling were deemed nice. The results of this study could impact social change by providing educators and caregivers with a deeper understanding of how girls experience self-image mediated by racial standards of beauty.

Doll play is a common activity for millions of young girls. Using a doll to act out everyday life during dramatic play allows children to learn many things: how to adopt more than one role at a time, how to display nurturing behavior, and how to problem solve. By the age of four, children are engaging in cooperative and pretend play.² Doll play is an important part of children's development. Taking care of dolls by dressing and feeding them improves fine motor skills, while role-playing offers the opportunity to develop language and social skills such as sharing and helping.² Pretend play also fosters the creative process and allows children to act out what they experience in the real world.²

Researchers have established that by the age of four, most children are aware of racial cues and can correctly identify skin color.³ In order to promote acceptance of diversity and healthy self-image, children in preschool and primary school would benefit from engaging in pretend play with dolls that more realistically display skin colors, hair, and facial features that represent the diversity of the community.² Ideally, as dolls become more complex in terms of

style, they will also become more ethnically diverse. Examples of these toys are fashion dolls —like Barbie— which are dolls that “place an emphasis on fashionable clothes and other possessions”.²

For years, Barbies have been a popular choice among young girls. Mattel estimates that 90% of U.S. girls between the ages of three and ten own at least one Barbie doll; girls between three and six own an average of twelve dolls each.⁴ The popularity of fashion dolls gives them the capability to affect the lives of young girls. Children are not blind to race and “studies have shown that the self-esteem of young girls is easily influenced even before they hit puberty”.¹ Because doll play has an important role on child development, fashion dolls should realistically represent different racial and ethnic backgrounds. In the United States, the dominant —or mainstream—culture is white culture. Its standard of beauty primarily consists of straight blonde hair, blue or green eyes, and thin lips and noses. It is evident everywhere; photographs of white women appear in fashion magazines, on billboards, and in advertisements at

department stores. What happens, then, when young girls do not have access to dolls that display the diverse range of skin colors, hair types, and facial features of various racial backgrounds? Limiting doll choices to just a few generic ones that do not deviate in appearance from a white, European standard of beauty can quickly teach children that one particular race is more valuable.

Statement of the Problem

Young girls are in a critical stage of development. This refers not only to the development of the language and social skills necessary to function in society, but also the development of racial attitudes that will contribute to their self-image and self-concept. Research has shown the important role of play in this developmental stage, including playing with dolls.³ Nevertheless, access to fashion dolls that accurately represent the varied skin colors, hair textures, and facial features of the girls is limited. Even when available, other factors such as price and popularity might be influencing girls' doll preferences, racial attitudes and self-worth.¹

Previous doll task studies have explored the relationship between doll preferences and racial attitudes in black preschoolers or white pre-schoolers.³ These studies presented participants with generic black dolls and white dolls that were not wholly accurate representations of the varied skin colors, hair textures, and facial features present in black African American or white European populations. Additionally, how children of other races, ages, and ethnicities respond to racial identity, racism, and racial awareness was not examined.

Still, this is an important area of research. There is evidence supporting the hypothesis that young children who have access to dolls that display the diverse range of skin colors, hair types, and facial features of various racial backgrounds are less likely to choose a doll that represents a white, European racial background.⁵ Children know the difference between dolls, not just because of their skin color. They assign values to dolls based on race.¹ By the age of four, children can recognize racial cues; if doll options are limited, perhaps a child's social development in terms of race and racial attitudes will be limited as well. For counselors, educators, and caregivers, it is essential to explore the perceptions that young girls have of dolls when mediated by race and racial features, in order to understand their personal experience, self-image, and self-worth.

Purpose of the Study

The purpose of this qualitative summative content analysis study was to explore the perceptions towards race and racial features that influence doll selection, and uncover the meaning girls attached to these features. The study was conducted among 26 girls, ages four to ten, attending a daycare, preschool, or elementary school in cities across southeast Texas. The recruitment process and interviews happened at the sites that agreed to host the study. The study was limited to female participants, since they more openly and commonly play with fashion dolls.

Guiding Research Questions

1. What are the perceptions that influence doll choice in girls aged four to ten?
2. Given access to a broad range of doll choices that display a diverse range of skin colors, hair types, and facial features of various racial backgrounds, how will girls aged four to ten describe their choices of dolls?
3. What meanings do girls attach to doll choice and dolls' racial characteristics?

Literature Review

Multiethnic (Jewish, Arab, Japanese, etc.) and Multiracial (Black, Asian, White, etc.) Barbie dolls with ambiguous, stereotypical facial features, hair textures, and skin colors are not adequate representations of how people actually look. As the United States increasingly becomes a more diverse country, there is a growing need for accommodations that reflect the reality of its citizens. Without a doubt, there are demands for diverse dolls that help foster positive racial awareness and racial identity in children. There is a relationship between doll choice and children's racial identities. Limiting children's doll choices to just a few generic ones that do not deviate in appearance from a white European standard of beauty can quickly teach children that their particular race is better or worse compared to others. The following six studies attempt to demonstrate and support the hypothesis that girls who have access to dolls that display the diverse range of skin colors, hair types, and facial features of various racial backgrounds, are less likely to choose a doll that is more socially acceptable (i.e. one that represents a white, European racial background).

Role of Doll Play

Doll play is an important part of young children's lives because acting out everyday life, via a doll during fantasy play, allows them to learn many things: how to adopt more than one role at a time, how to display nurturing behavior, and how to problem solve.² By the age of four, children are "playing cooperatively with peers" and "using dolls to act out family and school roles [and] roles they have observed, either real or fantasy".² Doll play is an important part of children's development, and taking care of dolls heightens fine motor skills. In addition, when children take on roles and play with other kids they "practice language and social skills, including sharing... and helping".² At this point in children's development, it becomes increasingly important for them to play with dolls that are diverse. If, for example, in school settings where there may be children who are all the same race or ethnicity, access to dolls that portray community diversity is still needed. Dolls should display diverse features that go beyond skin color: facial features, hair textures, and hairstyles must vary.

Black Barbie Dolls

Raynor addressed doll choice and how it can affect young girls' racial identity by analyzing the meaning that black Barbie dolls can hold for those who collect them. She admitted to "often wonder[ing] how [her] first black Barbie doll" contributed to

developing her identity as a "young black girl".⁸ Due to her appearance (brown skin, red hair, brown eyes), Raynor believed that she was unattractive and an outsider in her family. Her looks were never complimented because she did not meet the "black standard of beauty".⁸ She expressed that she did not ever want to look like a Barbie doll and that, at a young age, discovered the need to create her own beauty standards and accept her racial identity. It was posited that manufacturers took for granted the fact that young Black girls needed dolls "who reminded them that they are also beautiful and worth the time to emulate".⁸

Despite the negative opinions American society had about Barbie dolls, Raynor believed her first black Barbie doll represented positive concepts such as "identity, acceptance, and power".⁸ Dolls can help girls, who are members of a sex that is "already perceived as the Other", understand how the relationship between race, gender, and class is structured in society. Mattel tried many times to "broaden the ethnicities of their dolls, even though the focus was still" on light skinned blacks for the most part. Raynor stated it was necessary for young black girls to have dolls that were representative of the diverse range of skin colors present in African Americans.⁸

Overall, through disclosing a positive experience with having a black Barbie doll, Raynor revealed that there is a connection between doll choice and how children feel about their racial identities. Raynor's work supports the hypothesis that young girls who have access to dolls that display the diverse range of skin colors, hair types, and facial features of various racial backgrounds, are less likely to choose a doll that is more socially acceptable (i.e. one that represents a white, European racial background). She by revealed that having dolls is an important aspect of young girls' social development as well as racial identity, and that they should have access to dolls that represent the varied skin colors, hair textures, and facial features that are present in the race(s) of their market.⁸

Doll Selection and Racial Cues

A seminal piece of research supporting the current study was the 1974 work of Katz and Zalk. These researchers examined factors that influenced the development of racial attitudes in young children. Katz and Zalk found that children in preschool could acknowledge, and were familiar with, racial cues. By the time they have reached their fourth birthday, children can accurately recognize skin color as a racial cue.³

In the Katz and Zalk study, kindergartners and preschool children were given a doll task by same race and other race examiners. From

an assortment of black and white dolls, the children were asked to select their favorite doll, their least favorite doll, the good doll, the bad doll, the doll that was a nice color, and the doll that they wanted to take home with them.³ Following their doll selections, the children were also asked to indicate which doll they believed looked the most like them, and to point out the doll that looked like a white child or a black child. Overall, all children in the study chose same race dolls more often and associated them with positive characteristics while associating other race dolls with negative characteristics. The researchers also found a connection between the children choosing dolls and the examiner who administered the doll task. For example, the white kindergarten children typically chose white dolls if the examiner was white and black dolls if the examiner was black.

A weakness of this study is that it used only one racial cue, the skin tone of the dolls, to determine whether the doll was black or white. Katz and Zalk did not find the “strong preference for white dolls” found in previous studies.³ This was due in part, perhaps, because the male and female dolls did not differ in hair texture or color, eye color, skin color, or facial features. They were chosen because their “facial features were constant across skin color, and noses and mouths were judged...to be ambiguous enough to be either European or African”.³

If the study were to be replicated, in order to achieve responses that are more substantive, the black dolls should consist of several options that simulate the diverse range of skin colors, facial features, and hair textures that are present in African American populations. For example, a light skinned black preschooler who has medium length curly hair and hazel eyes may not identify with a generic, dark brown doll that has brown eyes and black hair. With the dark brown doll being the only offering under the “black doll” category, the child may indicate the doll as looking the most like them simply because there are no other options. As a result, a forced response (i.e. one that is more socially acceptable) is produced. A second limitation to the study is that Katz and Zalk only selected black children and white children to participate in the doll choice task. Their goal was to assess the “acquisition of racial attitudes in young children”.³ However, the phrase “young children” does not specify a racial or ethnic demographic. Children of other races and ethnicities deal with racial awareness, racial identity, and racism. Thirdly, the fact that examiner’s race was not constant is a limitation to the study.

Placing Value on Color

Highlighting the connection between doll choice and children’s racial attitudes, Perez discussed the racial attitudes of pre-school-aged black children. In another experiment, children were presented with two dolls, a dark skinned doll and a white doll.¹ They were asked which one they preferred and why. Most of the children stated they liked the white doll because it was the nicest one. When it came to the dark skinned doll, those same children described it as the bad one. In the rationale for her study, Perez recognized that young girls showed signs of self-hatred and “internalized racism”.¹ She also discussed the idea that the young girls preferred the white doll because modern society –like that in the U.S.—has bombarded people with the idea that white, European racial and ethnic backgrounds should be (and are) the most popular beauty standard. As evidence of this fact, African American Barbie Dolls tend to be cheaper than “otherwise identical” white Barbie dolls. Perez stated that “[By] devaluing the black doll...young black girls [are lead] to devaluing themselves...[by] internalizing that hatred”.¹

As shown by multiple studies, young children who have access to dolls with a limited range of skin colors, hair types, and facial features of various racial backgrounds are more likely to choose a doll that is more socially acceptable (i.e. one that represents a white, European racial background).^{1,3,5} From this, the following questions can be asked: “If children are presented with a more diverse range of doll choices, will they choose a doll more similar to their racial or ethnic backgrounds or will they choose the doll that is a more racially popular choice?”

The next section will examine the methodology of the proposed research.

Methodology

Qualitative Research Design

The present study was conducted under the qualitative research paradigm. Qualitative inquiry involves inductive data analysis from a purposeful sample, usually studied in natural settings.⁶ Qualitative research explores in-depth the experiences of the participants and the meaning they might derive from the process, while maintaining a contextually sensitive understanding of the topic being studied.⁹ For this qualitative study, the researchers used a summative content analysis methodology. Content analysis encompasses systematic techniques used to analyze the informational contents of textual data. Although it can also be used for quantitative designs, it is most commonly utilized as a qualitative technique.^{10,11} Summative content analysis requires the identification and counting of

keywords, their further comparisons, and the interpretation of their underlying and latent context.¹² In the present study, summative content analysis was conducted to study the perceptions of young girls when selecting fashion dolls with diverse racial and ethnic features, and the meaning derived from this process.

The data collection and analysis process consisted of recorded interviews that were analyzed for themes. Participants were presented with dolls that had a range of attributes and were asked a variety of questions related to doll choice. Each girl was asked the same set of interview questions by the same examiner. Their responses were tape-recorded and transcribed into an Excel spreadsheet for coding. The goal of summative content analysis is to uncover any themes, or factors, that determine the process used by the girls in making choices about doll selection.¹² In this study, the units of analysis were individual qualitative interviews with the participants and the units of meaning were the themes discovered after analyzing the interviews.

Participants and Context

Participants were 26 girls between the age of four and ten attending a daycare, preschool, or elementary school in southeast Texas. Recruitment occurred through the use of flyers at the sites that agreed to host the study. Informed consent was obtained from parents or legal guardians, and participants were given an opportunity to sign an informed assent if they were cognitively capable of doing so. In accordance with qualitative inquiry, the study used a purposeful sampling method on a volunteer basis.¹³ Participants were selected if they met the population requirements (female, four-to ten-years-old), signed a letter of assent (if cognitively appropriate to do so), and received consent from their parents or legal guardians. Thirty-five girls met criteria for inclusion and participated in the study.

Materials

The materials consisted of thirteen 11.5 inch Barbie dolls dressed in similar dresses of equal length. Four dolls were dressed in a yellow dress, four dolls were dressed in a pink dress, three dolls were dressed in a black dress, and two dolls were dressed in a blue dress. The dresses were short, and none of the dolls were wearing shoes. The dolls had flexible, moveable arms and could easily be made to sit and stand in various positions. The dolls spanned a variety of races (Asian, black, white) and ethnicities (Hispanic/Latino, Japanese). Hair texture and color, facial features, and skin tones varied. The

participants' responses were recorded using a tape recorder and a video camera.

Data Collection Procedures

Each participant's interview was conducted separately in an empty classroom or meeting room at a local school or church. Upon their arrival to the venue, the parents were given a parental consent form to read and sign. Some of the participants who were able to do so signed a letter of assent. The signed letters of assent and parental consent forms were collected and stored in a file folder, which were then placed into a sealable metal container that could be locked. For the interview, the participants were shown all of the dolls (which represented a variety of ethnically and racially diverse populations) and were asked questions about them. Following qualitative interview standard practices, the interviewer responses were limited and included only minimal encouragers (e.g. uhhuh, I see, tell me more) to elicit a full response set.

The interview questions were as follows:

Show me the doll that you like best or that you would like to play with. Why did you choose that one?
Show me the prettiest doll. Why is she the prettiest?
Show me the ugliest doll. Why is she the ugliest?
Show me the doll that looks most like you. How does she look like you? If you could take home one of these dolls today, which doll would you take home with you? Why is that? Point to the nice doll that helps her friends. Why did you choose her? Point to the bad doll that does not help her friends. Why did you choose her?

Role of the Researcher

In qualitative inquiry, the researcher is considered a variable in the study since the data obtained is analyzed through the lens of the researcher.⁶ Because the researcher becomes the primary instrument of data collection and analysis it is crucial to be aware of the influence this might have in the study.⁹ As a child, the primary researcher played with Barbies frequently. Whenever her grandmother or mother would bring her along with them to the grocery store, she would beg for a Barbie doll. By the time the lead researcher was 7, she had a rather large collection of Barbies including several versions of Barbie, Ken, Skipper, and Christie. Christie was Barbie's black companion, and by far the researcher's favorite Barbie

doll. She states "Christie wasn't a realistic representation of myself or any other black person I knew, but she did bring some much-needed diversity to my Barbie doll collection. Christie was very precious to me, and it seemed my mother liked her as well. Whenever I spent a few hours playing with Barbie dolls on the weekend or after school, my mother would always encourage me to include Christie in whatever scenario I came up with for that day. She wanted to show me that black people belong, that they are a part of [the] world's narrative" (S. Anderson-Bledsoe, personal communication, January 25, 2016). From an early age, the primary researcher knew that racial and ethnic representation was important, and that there was a relationship between racial and ethnic representation and self-esteem. Although too young to put it into words, she knew that developing a healthy racial identity had something to do with the toys she played with. This knowledge resulted in an interest in researching the topic of race and doll selection. Because of the potential for bias, additional researchers were brought into the project to develop the interview process and conduct data analysis.

Trustworthiness

Trustworthiness was achieved by considering the appropriateness of the research questions and participant sample and by utilizing the appropriate methods for data collection and analysis.⁶ This study used the methods of summative content analysis to obtain triangulation, including inter-coder agreement.¹² The researcher was trained in qualitative interview protocol prior to conducting the interviews. After the interviews took place, the researcher's mentor later reviewed the transcripts and themes.

Data Analysis

Upon completion of the interviews, the interviews were transcribed. Background noise or technical difficulties with the recording made nine of the interviews unusable. Thus, 26 transcripts were available for analysis. Transcription data was placed in an Excel spreadsheet. The most objective part of a qualitative summative content analysis is based on a simple word count for each question.¹² From this word count, data were clustered based on keywords in context for each of the questions.¹⁴ Responses were grouped into emergent themes and coded according to major focus or intent of the statement. Next, the number of occurrences per code were counted and interpreted. Finally, a summative content analysis was used to analyze the questions.¹² The source for naming of categories was investigative and the verification component

was technical (e.g., use of inter-coder agreement).¹⁵ Specifically, the two researchers independently coded 20% of the interview data and after establishing 100% interrater reliability, using Cohen's Kappa measure, then coded the remaining data together.¹⁶ To integrate responses across questions, each of the participants' data were subjected to cross case analyses to determine whether emergent themes could be disaggregated.¹⁷

Results

Several findings emerged as the result of the coding of interview data. They will be discussed here in relation to the questions asked to each participant.

Question 1: Show me the doll that you like best or that you would like to play with. Why did you choose that one?

Participants selected doll #5 most frequently, with six girls choosing it as the one they liked best. Dolls #6, #9, and #12 were each selected three times. The top keywords relevant to why they chose the dolls they selected were "pretty/beautiful" (n=16), "looks like me" (n=16), "hair" (n=14), "brown" (n=9) and "eyes" (n=8). Hair, overall attractiveness, eye color, and the participant's perception of the doll's features being similar to theirs were themes that emerged when asked which dolls the participants liked best and wanted to play with. Finally, nine of the participants referenced a female role model that reminded them of the doll that they chose.

Question 2: Show me the prettiest doll. Why is she the prettiest?

Responses varied to this question with dolls #5 and #9 being selected four times each and dolls #3, #7, #8, and #12 being

chosen three times each. The top keywords relevant to why they chose the dolls they selected were "pretty/beautiful" (n=16), "hair" (n=11), "eyes" (n=6), "face" (n=5), and "different" (n=5). Overall attractiveness, hair, eyes, and face were themes that emerged when asked which dolls were the prettiest. Finally, six of the participants mentioned they chose a doll because of its uniqueness.

Question 3: Show me the ugliest doll. Why is she the ugliest?

Unlike the prettiest doll, the ugliest doll responses were largely focused on two dolls: doll #1 (n=8) and doll #9 (n=5). Keywords relevant to why these dolls were the ugliest were "hair" (n=18), "skin" (n=8), and "dark" (n=6). Overall unattractiveness, face, plainness, hair, and skin were themes that emerged when asked which doll was the ugliest. Hair that is puffy or poofy and

skin that is dark appeared to make dolls unattractive, as did plainness. Uniqueness or specialness, as found in the Q2 results, made a doll attractive while plainness or boringness made a doll unattractive. Also, four of the participants chose a doll as the ugliest because it was weird or fake looking. Dolls that were not realistic in appearance were deemed unattractive.

Question 4: Show me the doll that looks most like you. How does she look like you?

The top keywords relevant to why the participants chose the dolls they selected were “hair” (n=25), “brown” (n=16), “color” (n=14), “skin” (n=13), “eyes” (n=10), “dark” (n=6), “light” (n=6), “looks like me” (n=5), “face” (n=4), and “nose” (n=4). The participants’ responses were mainly focused on doll #5 (n=6) and dolls #1, #6, and #9 (n=3). It is important to discuss which participants were or were not accurate in choosing which doll actually looked like them. Out of the 26 participants, only three incorrectly chose which doll actually looked like them. Two of those three participants were eight-years-old and one was six-years-old. One of the eight-year-olds was biracial (black and white) and chose doll #5 as looking most like her. The other eight-year-old was black and chose doll #5, and the six-year-old was black and chose doll #9. Dolls #5 and #9 were two of the most popular dolls and they were chosen incorrectly for Q4 three times.

Question 5: If you could take home one of these dolls today, which doll would you take home with you? Why is that?

The top keywords relevant to why the participants chose the dolls they selected were “looks like me” (n=17), “prettiest” (n=9), “black” (n=6), “hair” (n=6), and “mom” (n=4). The dolls that were chosen most often were doll #5 (n=6), doll #4 (n=3), doll #6 (n=3), and doll #9 (n=3). Interestingly, in response to Q2, participants chose a doll as the prettiest if it was different, unique, or perfect in appearance but in Q5, 17 out of 26 participants chose a doll that looked like them as the one they would want to take home.

Question 6: Point to the nice doll that helps her friends. Why did you choose her?

The top keywords relevant to why the participants chose the dolls they selected were “looks like” (n=12), “nice” (n=16), “smart” (n=3), “smile/smiling” (n=6), “good” (n=5), and “hair” (n=4). The dolls chosen the most in response to Q6 were dolls #3 and #11 (n=4) and doll #8 (n=3). Interestingly, there was no mention of skin or eyes and there was hardly any mention

of hair. Skin, eyes, and hair were mentioned several times in every question but Q6. It is evident that “helpful” is not a visual, even though Q6 asks for a visual. As discovered in Q2, prettiness is based on the appearance of the doll’s skin, hair, and eyes, while niceness is based on whether or not the doll was smiling.

Question 7: Point to the bad doll that does not help her friends. Why did you choose her?

The top keywords relevant to why the participants chose the dolls they selected were “mean” (n=10), face (n=5), “diva/fancy/stuck-up” (n=5), “not/isn’t smiling” (n=4), “eyes” (n=4), “hair” (n=4), “grouchy” (n=2), and “bossy” (n=2). Doll #1 was chosen nine times, doll #7 was chosen six times, and doll #2 was chosen five times. Interestingly, in Q6 “helpful” is not a visual but in Q7 “mean” is a visual. Lastly, three participants mentioned an external female that behaved in a negative way that reminded them of the doll they chose. The negative external females were “diva”, “cousin”, “actress”, and “mom”.

Discussion

The themes that follow are those that were revealed in the coding process. Each theme will be discussed individually.

Themes

Goldilocks to grandma.

In Q1, nine participants mentioned an external female role model that reminded them in some way of the doll they chose. The doll’s similarity in appearance to an external role model appears to be one of the factors that determined which dolls the participants liked best and wanted to play with. The external female role models they mentioned were goldilocks, sister, princess, model, teacher, mommy, diva, Selena Gomez, Demi Lovato, and grandma.

Actress to angel.

In Q6, seven participants mentioned an external female role model that reminded them in some way of the doll they chose. The doll’s similarity in appearance to an external role model appears to be one of the factors that determined which dolls the participants thought were nice. The external female role models they mentioned were actress, mom, teacher, sister, and angel.

Looks like me.

The doll’s similarity in appearance to the participants was one of the factors that determined doll choice. In response to Q1, nine participants mentioned that they chose the doll because it looked like them. In response to Q5, 17 participants mentioned

that they chose a doll because it looked like them. In response to Q6, 12 participants mentioned that they chose a doll because it looked like them. “Looks like me” arose several times in each question except for Q7 and Q3. This finding supports the idea that when girls have access to a broad range of doll choices, they tend to prefer a doll more similar to their own ethnic or racial background.

Hair, skin, and eyes.

Hair, skin, and eyes were three of the most common factors that determined doll choice for most of the questions (whether the hair, skin, and eyes of each doll was similar in appearance to an external female role model that the participant looked up to or the hair, skin, and eyes of each doll reminded the participant of themselves). Only in Q6 was there was no mention of skin or eyes and very little mention of hair. In Q6, participants’ concept of niceness was related to whether the doll was smiling or not rather than how pretty the doll was. Therefore, the prettiest dolls were not necessarily the nicest dolls and the ugliest doll were not necessarily the meanest dolls.

Female role model/external female influence.

The dolls’ similarity in appearance to an external role model appears to be one of the factors that determined which dolls the participants preferred. In Q1, nine participants mentioned an external female role model that reminded them in some way of the doll they chose. In Q6, seven participants mentioned an external female role model that reminded them in some way of the doll they chose. In Q7, three participants mentioned a negative external female that reminded them of the doll they chose. Doll #1, doll #5 (n=2), doll #6 (n=2), doll #8, doll #9 (n=2), doll #10, and doll #12 were chosen in response to Q1 when participants mentioned an external female role model. Interestingly, the most popular (most liked) dolls were dolls #5, #6, and #9. Doll #1, doll #3 (n=3), doll #6, doll #11, and doll #12 were chosen in response to Q6 when participants mentioned an external female role model. Lastly, dolls #4, #12, and #11 were chosen in response to Q7 when participants mentioned an external female role model.

Doll #1.

One of the least popular dolls (most disliked) was doll #1. It was chosen the most in response to Q3 (n=8) and Q7 (n=9). It was chosen very little in response to Q5 (n=1), Q2 (n=1), Q1 (n=1), and Q4 (n=3). It wasn’t chosen at all in response to Q6. In Q7, posture seems to have played a role in some of the participants’ doll choice. Three

out of 26 participants mentioned that they thought doll #1 was a bad doll because one of her hands was on her hip. Interestingly, dolls #2, 3, 4, and 13 all had one of their hands on their hips. Dolls #2, 3, and 13 were white and doll #4 was Biracial (black and white). Doll #1 was a dark-skinned black doll. Thus, it appears that Doll #1 was one of the least popular dolls and was perceived in Q7 as having a bad attitude because of her skin color. The perception of a doll having a bad attitude coincided with her hand being on her hip only if her skin is dark.

Nonrealistic dolls are ugly.

In response to Q3, four participants chose a doll as the ugliest because it was weird or fake looking. Dolls that were not realistic in appearance were deemed unattractive. Doll #4 (n=2), doll #1, and doll #13 were chosen in response to Q3 when participants mentioned weirdness or fakeness. Interestingly, one of the participants that described doll #4 as unrealistic was a fouryearold. This supports the idea that by the age of four, it becomes increasingly important for children to play with dolls that are realistic.² Dolls that do not realistically represent a certain racial or ethnic group may be rejected by young girls and looked upon unfavorably.

Smiling.

In Q6, one of the top keywords relevant to why the participants chose the dolls they selected was “smile/smiling” (n=6). Doll #3 (n=2), doll #7, doll #11 (n=2), and doll #13 were chosen in Q6 as the good dolls because they were smiling. Upon examining themes and keywords (word repetitions), it was found that overall prettiness or attractiveness is based on skin, hair, and eyes while niceness is based on smiling. This is interesting because it contradicts with the commonly held belief that someone who has a happy, inviting smile is considered to be attractive.

Implications

There are a number of implications for this research. First, a better understanding of how young girls perceive beauty, and themselves, is essential to child development studies. Therapists and educators working with girls can use these results to better understand the population. Secondly, the concept of beauty is strongly linked to the appearance of skin, eyes, and hair. Since only hair has the ability to be altered, one’s own sense of beauty may be rooted in genetics. If young girls have a positive experience with having a doll that reflects their racial or ethnic background, they would be one step closer to developing a healthy identity. Finally, these girls looked to external female

role models to set the standard for beauty and desirability. Adult women who are cognizant of this, particularly mothers, teachers, and other women of authority, can positively influence the development of beauty and desirability in girls.

Further Research

It would be helpful to see how other populations or sample ages perceive doll choice and whether findings are similar or different from those in the current study. Further research could include boys age four to ten or explore the perceptions of girls aged eleven to seventeen who no longer play with dolls. Additional research may include direct observation of children as they make doll selection. Direct observation removes the interviewer from the study and has the potential to reduce bias.

Conclusion

Originally, we hypothesized that young children who have access to dolls that display the diverse range of skin colors, hair types, and facial features of various racial backgrounds are less likely to choose a doll that is more socially acceptable (i.e. one that represents a white, European racial background). Findings related to race and ethnicity portrayed by the doll were anticipated. However, this study’s results reveal that it is not the ethnic makeup of the doll that is important but how the girl sees the doll in relation to herself. This is a powerful finding that can impact psychology and education. The use of qualitative interviews helped reveal themes that transcend the mere choosing of a doll, and help to explain the question “why”. The results of this study showed the importance of having a doll that “looks like me,” as well as the doll’s connection to an external female role model. Future research is necessary to continue to explore the factors influencing doll selection in girls, and moreover, the development of their racial attitudes.

Doll Characteristics

Doll Number	Dress Color	Skin Color	Race or Ethnicity	Hair Color	Hairstyle	Eye Color
#1	Pink and White with Black Dots	Dark Brown	Black	Black with Brown Highlights	Ponytail	Light Brown
#2	Hot Pink	Tan	White	Strawberry Blonde	Ponytail	Light Brown
#3	Hot Pink	White	White	Black with Brown Highlights	Ponytail	Light Brown
#4	Yellow and White	White	Biracial; White and Black	Blonde	Ponytail	Green
#5	Pink and White with Black Dots	Medium Brown	Black	Light Brown	Ponytail	Light Brown
#6	Yellow and White	White	White	Blonde	Ponytail	Blue
#7	Yellow and White	Pale White	Asian (Japanese)	Black	Ponytail	Light Brown
#8	Blue and Pink	Tan	Latino	Black with Brown Highlights	Ponytail	Dark Brown
#9	Blue and Pink	Light Brown	Black	Black	Ponytail	Dark Brown
#10	Yellow and White	Tan	White	Dark Brown	Ponytail	Green
#11	Hot Pink	White	White	Honey Blonde	Ponytail	Light Brown
#12	Hot Pink	Brown	Black	Brown	Ponytail	Light Brown
#13	Pink and White with Black Dots	Tan	White	Reddish Brown	Ponytail	Green

Demographics and Doll Choice of Participants

Participant	Age	Race or Ethnicity	Doll Choice for Q1	Doll Choice for Q2	Doll Choice for Q3	Doll Choice for Q4	Doll Choice for Q5	Doll Choice for Q6	Doll Choice for Q7
#1	10	Biracial; Black and White	9	9	8	9	9	10	7
#2	6	White	6	3	2	11	3	9	10
#3	9	Black	12	12	9	12	12	11	1
#4	7	White	3	3	9	6	3	8	7
#5	10	Black	5	7	13	1	1	1	2
#6	8	Biracial; Black and White	5	12	1 and 9	5	5	4	1
#7	9	White	11	3	2	10	8	13	1
#8	9	Biracial; Black and Latino	5	8	9	5	5 and 8	3	1 and 7
#9	10	White	13	10	9	6	5	11	7
#10	4	Black	5	5	1	5	5 and 13	13	4
#11	4	Black	9	9	4	9	5 and 9	12	2
#12	8	Biracial; Black and Latino	1	1	6	5	5	5 and 8	7
#13	9	Biracial; Black and White	4	12	1	4	4	2	1
#14	10	Black	12	9	7	1	12	12	2
#15	6	Biracial; Black and White	9	9	1	4	4	6	2
#16	6	Black	12	5	11	9	9	3	11
#17	10	Latino	8	8	1	8	8	3	12
#18	10	Latino	8	8	1	5	5	9	2
#19	7	Asian (Indian)	5	5	6	1	1	6	11
#20	9	Asian (Indian)	5	5	1	5	5	7	2
#21	10	Biracial; Asian (Japanese) and White	7	7	6	7	7	7	1
#22	4	Biracial; Asian (Japanese) and White	7	7	4	7	7	3	13
#23	4	White	6	6	1	6	6	8	1
#24	8	White	10	10	12	13	13	10	1
#25	9	Biracial; Black and Latino	4	4	7	4	4	11	6
#26	10	White	6	6	13	13	13	11	7

References

- ¹Perez, A. (2012) "How Racism Affects Children: The Doll Test." *Feminspire*. <<http://feminspire.com>> (Accessed 10/15/2015)
- ²Texas Child Care (2004) "Dolls and doll play: A new look at a familiar prop." *Texas Child Care Quarterly*. Summer 2004.
- ³Katz, P.A., & Zalk, S.R. (1974) "Doll Preferences: An Index of Racial Attitudes." *Journal of Educational Psychology*. 66.5. Pg 663-668.
- ⁴Griffin, J. (2009) "Academics like to play with Barbies, too." *Pacific Standard*. <<http://www.psmag.com>> (Accessed February/20/2015)
- ⁵London, B. (2015) "Nigerian doll created by man who couldn't find a black toy for his niece is so popular in his country that it's outselling BARBIE." *The Daily Mail*. <<http://www.dailymail.co.uk>> (Accessed 04/05/2015)
- ⁶Creswell, J. W. (2014) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Los Angeles: SAGE Publishing.
- ⁷Porter, J. R. (1971) *Black Child, White Child; The Development of Racial Attitudes*. Harvard University Press.
- ⁸Raynor, S. (2009) "My first black Barbie: Transforming the Image." *Cultural Studies/Critical Methodologies*. 9.2. Pg 179-185.
- ⁹Patton, M. Q. (2015) *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*. SAGE Publications.
- ¹⁰Forman, J. & Damschroder, L. (2008) "Qualitative Content Analysis." *In Empirical Methods for Bioethics: A Primer*. Jacoby, L., and Siminoff, L.A., eds. Emerald Group Publishing Limited. Pg 39-62
- ¹¹Alok, S., Raveendran, J., & Shaheen, M. (2014) "Conflict Management Strategies Used by Indian Software Companies: A Summative Content Analysis." *IUP Journal of Soft Skills*. 8.4. Pg 47-61.
- ¹²Hsieh, H. F. & Shannon, S. E. (2005) "Three Approaches to Qualitative Content Analysis." *Qualitative Health Research*. 15.9. Pg 1277-1288.
- ¹³Jupp, V. (2006) *The SAGE Dictionary of Social Research Methods*. SAGE.
- ¹⁴Fielding N. G., & Lee, R. M. (1998) *Computer Analysis and Qualitative Research*. Thousand Oaks: SAGE Publishing.
- ¹⁵Constas, M. A. (1992) "Qualitative analysis as a public event: The documentation of category development procedures." *American Educational Research Journal*. 29. Pg 253-266.
- ¹⁶Siegel, S., & Castellan, J. N. (1988) *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill.
- ¹⁷Miles, M. B., Huberman, A. M., & Saldaña, J. (2014) *Qualitative Data Analysis: A Methods Sourcebook*. SAGE Publications.

Agrarian Capitalism and the Privatization of Collective Land in Morocco

BY DAVID BALGLEY

UNIVERSITY OF PUGET SOUND

Abstract

The recent approval of Morocco's Employability and Land Compact – a bilateral agreement between the Government of Morocco and the Millennium Challenge Corporation – will convert collectively-managed land in the Gharb irrigated perimeter into private property. This article seeks to address how this new program is likely to affect contemporary agrarian structures in Morocco, and how these shifts will impact rights-holders to collective land as their land is converted into private property. I assert that this project is based upon a particular interpretation of neoliberal ideology that has promoted a capitalized, export-oriented agricultural sector at the expense of small farmers in the traditional sector. This ideology dates back to the colonial era, and has been continued by the current regime in order to forge ties with a class of agrarian capitalists and rural elites by providing economic benefits in exchange for political support. Furthermore, I argue that converting collective land into private property will strengthen agrarian capitalist structures, encourage concentration of land ownership, and enhance processes of proletarianization affecting rural populations, which will in turn increase pressures to diversify livelihood strategies beyond small-scale farming.

Introduction

Land ownership plays a significant role in the distribution of political, economic, and social power. In turn, land reform processes are shaped by existing power dynamics and hierarchies. This makes tenure structures, systems of land distribution, and the design of land reform a vital area of inquiry when looking at the transformation of agrarian systems. There has been considerable contention among scholars and development theorists on how land tenure should be reformed and how these changes impact different population groups. This article—based on a review of theoretical and context-specific literature and supplemented by fieldwork in June and July of 2015—assesses how Morocco's Land and Employability Compact with the Millennium Challenge Corporation is likely to affect contemporary agrarian structures in Morocco, and how these shifts will impact rights-holders to collective land as their land is converted into private property.

The Compact includes a project that will convert collectively-managed land within the irrigated Gharb perimeter in northwest Morocco into private property. This pilot project will initially only impact approximately 3% of Morocco's collective land, but will eventually be scaled up throughout the country. It is part of a larger series of neoliberal policies and programs designed by the Moroccan regime that have had a powerful role in creating an agrarian capitalist system.¹ Since gaining independence, the government has promoted agricultural development policies that prioritize irrigated, export-

oriented large-scale farming operations, largely to the benefit of a class of rural notables and agrarian capitalists, whilst small farmers and the traditional farming sector have been largely ignored.² These policies of marketization and liberalization have provided economic benefits to elites that the regime depends upon for political support.³

In this article, I argue that by opening irrigated collective land to conventional land markets through privatization processes, the Compact will strengthen agrarian capitalist structures, promote land concentration, and encourage rural livelihood diversification. Agrarian elites who hold large amounts of liquid capital will be able to expand their holdings as legal barriers to purchasing irrigated collective land are eliminated. At the same time, rights-holders to collective land will be able to sell or rent the land to which they gain a title, as their land assets become unfrozen. This will allow them to diversify their livelihood strategies beyond agricultural pursuits. These livelihood shifts have been observed in other areas of Morocco, particularly from pastoralism to a combination of cultivation, wage labor, and migration abroad.⁴ Concentration of land ownership and the pre-eminence of large operations in the agricultural sector will make small farming increasingly untenable as a single livelihood option for rural populations.^{5,6} This could lead to the semi-proletarianization of small farmers, as wage labor takes an increasingly important position in the diverse portfolio of livelihood strategies that rural households are deciding upon.⁷

Colonial influence on the development of agrarian structures and policies

The French Protectorate, which lasted from 1912 to 1956, had a profound impact on the development of agrarian structures and policies in Morocco. In many ways, contemporary agricultural policies that promote mechanization, modernization, and privatization arose out of colonial policies. France developed a bifurcated agricultural system through expropriation of land, created a new class of rural notables who largely collaborated with the regime, developed hydro-projects and irrigated perimeters, and instigated the *Dahir* of 1919, a royal decree which continues to shape the administration of collective land to this day.

French agricultural policy in Morocco first encouraged production of wheat, and later high-value fruits and vegetables, to be exported to France.² In order to fulfill this policy, French colonists expropriated some of the most productive agricultural land in Morocco—approximately 1 million hectares, or 13% of Morocco's arable land.⁸ At the time of independence, 94% of this land was owned by only 2,600 European large landowners, and the average parcel size was 365 hectares.²

French colonists were not the only ones who expropriated land and created large landholdings. A new class of Moroccan rural elites also arose during the colonial period. Working closely with the French, they amassed approximately 1.6 million hectares of land with an average parcel size of 225 hectares.^{2,8} These rural elites occupied regional political positions in the colonial regime, and were often local notables who

used their power to accumulate land and other capital assets.^{9,10} While some of these individuals were stripped of their positions and land after independence, many of them continued to be powerful actors in rural areas.^{11,12}

Another notable feature of the latter part of the colonial project was the emphasis on agricultural development through irrigation projects. Officials expected that investments made in the most productive land would generate taxes and rents that could pay off the cost of new infrastructure.² This policy tended to benefit large landowners, both Moroccans and Europeans, who already controlled the most productive agricultural land.¹³ The colonial regime created the Office de l'Irrigation in 1941 and developed a plan to perennially irrigate 1 million hectares of agricultural land by the year 2000. This institution pioneered a development model in irrigated areas that continues to this day, characterized by centralized agricultural decision-making in government agencies and *remembrement*.² *Remembrement* is the process of dividing and reallocating irrigated land into a geometric landscape appropriate for capitalized irrigated agriculture. Decrees in 1951 and 1953 opened land within the irrigated perimeters to speculation, allowing for the development of further irrigated large-scale farming operations.²

The vast majority of Moroccans remained in the traditional sector during the colonial period, engaging in subsistence production on small, non-contiguous parcels of communally-owned land.¹⁴ As land ownership became more concentrated, traditional landholdings shrank and rural peasants were pushed onto marginal land, which in the past had only been used for pastoralism.⁸ At the time of independence, over 60% of Moroccan land was owned by only 5-10% of Moroccan landowners, while 40% of the population had virtually no land.² This combination of land concentration and the growth of landless peasants contributed to the proletarianization of rural peasants and served as a basis for a new agrarian capitalist system.⁶

Agrarian capitalism is a mode of production distinguished by private ownership of the means of production for the purpose of selling agricultural commodities in a market system.¹⁵ The salient features of capitalist farms are the utilization of wage labor, integration with national and international markets, and for-profit production.¹⁶ Agrarian capitalists are the landed elite who controls the means of agricultural production, and they are frequently absentee landowners who live in urban areas while occupying a powerful role in rural power structures.¹³ This system contrasts with the traditional

sector in Morocco, which is characterized by small, non-contiguous parcels totaling 4 to 10 hectares, subsistence production using family labor, and communal ownership.⁸ Agrarian capitalism requires a pool of agricultural laborers—often peasants who are either supplementing their income or have lost their land.¹³

In order to address rural discontent arising from land expropriation, the colonial regime decided to delimit and insulate collective tribal land from land markets.¹⁷ To this end, the *Dahir* of 1919 was created – a royal decree that gave perpetual usage rights to the ethnic collective that inhabited a particular area of land at the time of the *Dahir*.¹⁸ This decree protected collective land from expropriation as it made it inalienable—that is, it was jointly owned by the collectivity as a whole and could not be sold to outsiders, with the exception of the state in the form of eminent domain policies.⁹

Post-colonial agricultural development and the current situation of land

Many of the characteristics of the agricultural policies and systems of production created during the Protectorate period have continued in the post-colonial era. Morocco continues to exhibit a bifurcated agricultural system divided between the traditional sector and large-scale capitalized operations, and agrarian elites persist as a significant interest group in rural power structures. In addition, the government has largely sustained policies that promote the development of irrigated areas, although the details of these programs have changed since the colonial period.¹ This is seen by King Hassan II's announcement on January 8, 1969 that irrigating 1 million hectares was a national goal – an explicit continuation of colonial policies emphasizing irrigation development.¹⁹

Between 1956 and 1969, there was a series of administrative changes related to the management and development of irrigated areas, which largely continued colonial policies of centralized decision-making and state support for farmers in irrigated areas through the creation of cooperatives and contract-based farming.² This resulted in the contemporary system, which divides each of the irrigated regions into separate administrative zones that are managed by an *Office Régionale de Mise en Valeur Agricole* (ORMVA) [Regional Office for Agricultural Development]. Currently, there are 9 ORMVA that are ostensibly under the administrative umbrella of the Ministry of Agriculture, although in practice they are largely autonomous due to their financial independence.^{2,20}

As aforementioned, collective land in

Morocco is governed by the *Dahir* of 1919, which dates back to the colonial period. However, with regard to collective land within irrigated perimeters, a major evolution of land governance occurred in 1969 in the form of a new *Code des Investissements Agricoles* [Agricultural Investment Code] and an accompanying law that was adopted in order to convert collective land in irrigation schemes into private property.²⁰ This was partially influenced by a World Bank report that recommended investment in irrigation infrastructure.²¹ This report also suggested measures to recoup excessive gains on government investment by large privately-owned farms; although this recommendation was echoed by several Moroccan government agencies, no such measures materialized.² The combination of national and international pressures resulted in the 1969 Code, which has guided Morocco's agricultural development policy through the latter half of the 20th century.

This new Agricultural Investment Code and its associated legal reforms had two main aspects. The first of these aimed to focus investment in the irrigated perimeters managed by the ORMVA. These agencies continued centralized production policies and were responsible for choosing crops and rotation cycles, developing agricultural and water usage techniques, and building infrastructure within their jurisdiction.²² The second main aspect of the reforms associated with the 1969 Code was a modest land reform project, which included the conversion of collective land into private property.²

The new law in 1969 that allowed for the conversion of collective land into private property reformed the *Dahir* of 1919, but was only applicable in irrigated perimeters. The conversion process began with the establishment of lists of rights-holders to a given ethnic collectivity's land, which was carried out asymmetrically based on customary eligibility criteria that was subject to internal power dynamics.²³ This was followed by subdividing and individualizing plots based on *remembrement* procedures. However, implementation of this privatization procedure did not yield the expected results. Since 1969, no more than half of the collective areas have been delimited, and only a few were subdivided with individualized plots assigned to collectivity members. No comprehensive titling programs took place—by one account, only 46 titles were given out over the entire process.²⁰ The Government of Morocco identifies the major difficulties as: legal and technical issues related to the lack of clarity of the legal situation of collective land; lack of funds to implement division of lands; issues of coordination between government

agencies and ministries that do not have clearly defined roles; and problems resolving inheritance and land disputes, due to issues with establishing lists of rights-holders in accordance with local customary law.²⁰

In 1999 and 2014 the government held a series of regional and national dialogues related to tenure shifts and general land issues with a variety of stakeholders, including local authorities, parliamentarians, public administrations, and representatives of ethnic collectivities. Two main points of view arose out of the 1994 discussions: the first recommended privatizing arable collective land used for agriculture and maintaining collective administration of rangeland and forests, while the second advocated for reforming the management and valuation of collective land through cooperatives and partnering with private investors, but rejected tenure conversion processes. None of the recommendations were implemented, resulting in increased disputes over land and a general worsening of the situation due to population growth and land scarcity.²⁰

A second round of dialogues was carried out in five regions of Morocco in 2014, and the majority view aligned with the first viewpoint of the 1994-95 discussions: privatizing agricultural collective land while maintaining the collective status of rangelands and forests. Furthermore, stakeholders agreed that peri-urban collective land should be managed as part of urban policy, and that changes to the administration of collective lands must be both inclusive and participatory. These dialogues were followed by a national conference on land in December 2015 marked by the publication and reading of a royal letter, which provided guidelines for the development of a national land strategy and action plan as well as the development of recommendations for reforming Morocco's land system. These recommendations included revising legal frameworks related to land, improving governance tools, reforming the land tax system, and improving land financing mechanisms.²⁰

While the privatization of all arable collective land continues to face legal barriers to conversion as a result of the *Dahir* of 1919, the 1969 reforms provide the government with a legal basis to begin privatization processes within irrigated areas.¹⁷ In essence, the recent Employability and Land Compact is designed to overcome problems that previous privatization projects faced while incorporating the recommendations of the regional dialogues and conferences. It will also serve as a pilot project that can demonstrate the economic benefits of tenure reforms while larger institutional structures are being reformed.

Contemporary land tenure institutions in Morocco

Today, Morocco divides land into five categories, which exhibit administrative plurality.²⁴ The two most prominent tenure systems for agricultural land are private property (*melk*) and collective tribal land held in trust by the state (*soulaliyate*). The other categories of land are religious land endowments (*habous*), pre-Protectorate grants to military supporters of the Sultan (*guiche*), and state-owned land, but these tenure institutions do not play a significant role in contemporary agriculture. *Soulaliyate* land, or collective land, is communally-owned by what the French Protectorate called "ethnic collectivities". This remains the official term for a tribal confederacy, although a clear definition of its meaning was never fully developed.²⁵ It is important to recognize that an ethnic collectivity is a kinship group that holds usufruct rights to *soulaliyate* land and can be either Arab or Amazigh, the two dominant ethnic groups in Morocco. This paper focuses on *melk* and collective land, as together they make up more than 93% of all farmland in Morocco (see Figure 1).

partially explains the official registration policy. In interviews, both urban and rural landowners emphasized that they preferred *melk* as a tenure system, as it granted them the most control over decisions related to their land. *Melk* land can be sold or rented through market systems, and is inherited by the owner's descendants without interference from outside institutions.

In contrast, collective land is officially managed by the Ministry of the Interior on behalf of ethnic collectivities under the *Dahir* of 1919. The *Dahir* states that *soulaliyate* land is to be governed in accordance with local customary law.⁹ As a result, *soulaliyate* land exhibits highly asymmetric forms of administration that can vary widely, even within the same region.²⁸ Collective land makes up between 34% and 42% of total land in Morocco, although only approximately one-tenth of that is arable farmland.²⁴ There are 4,600 recognized tribal confederacies in Morocco, each of which holds usufruct rights to their traditional lands.²⁹ These lands are administered by the *Direction des Affaires Rurales* (DAR) [Directorate of Rural Affairs], a department within the Ministry of the Interior.

Figure 1: Farmland by tenure status²⁶

Tenure Status	Hectares	% of Total
Private Property (<i>Melk</i>)	6,618,130	75.8%
Collective Land (<i>Soulaliyate</i>)	1,544,656	17.7%
<i>Guiche</i>	240,441	2.75%
State Land	270,153	3.09%
Endowments (<i>Habous</i>)	58,843	0.66%
Total	8,732,223	100%

One official at the *Agence Nationale de la Conservation Foncière du Cadastre et de la Cartographie* (ANCFCC) [National Agency of Land Cadastre and Cartography], a public institution jointly funded by the Ministry of Finance and the Ministry of Agriculture, reported that the official policy of the government is to privatize and title all land through registration with the ANCFCC.²⁷ However, legal codes and historical *dahir* prohibit the other four types of land from being converted into *melk*, with few exceptions.²⁸ This remains a major policy issue that needs to be addressed, and the Compact positions itself as the first step in a new registration process.

Melk is the sole category of land that is subject to official tax policy.²⁷ By increasing the amount of *melk* land, the government is concomitantly broadening its tax base, which

Each ethnic collectivity elects a *naib* (pl: *nouab*), a representative responsible for all contact with the DAR and who manages the land of their collectivity in accordance with customary law.³⁰ *Soulaliyates*, or those who hold rights to collective land, reported that the *naib* was actually selected by a local council, rather than by an election within the ethnic collectivity. This council is largely made up of older men who exclude women and youth from decision-making power; institutional channels to replace *naib* do exist, but local power dynamics and social considerations often prevent these channels from being used.³⁰ This demonstrates the inequitable access to decision-making power that is the status-quo within ethnic collectivities, which allows council members and *nouab* to use their own interpretation

of customary law for personal economic benefit.³¹

From my discussions with *soulaliyates* in the Middle Atlas, I found that there are two contemporary forms of managing collective land: delimitation of land based on historical usage and community discussions, and keeping land undivided for pastoralism or to lease out. The DAR supports formal delimitation and partitioning of land, and has in some cases provided boundary documents to rights-holders, although these documents are not equivalent to a land title.^{9,29} Delimitation makes collective land individualized and inheritable, although it can still only be sold within the collectivity through informal markets. The Ministry of the Interior encourages the rental of *soulaliyate* land, subject to the approval of the *naib* and Ministry officials through an opaque process that lacks accountability.¹⁷ The Ministry collects the rent generated by lease agreements, which is put into a fund to be used either in individual disbursements to those included on the lists of rights-holders or to fund a development project for the benefit of all *soulaliyates*.²⁹ The combination of centralized bureaucratic control and local administration aligning with customary law contributes to a highly complex system of tenure practice that is often unintelligible to outsiders and incompatible with conventional land markets.³²

Morocco's Employability and Land Compact

Morocco's Employability and Land Compact (hereafter referred to as "the Compact") is a bilateral agreement between the U.S. Government, acting through the Millennium Challenge Corporation (MCC), and the Kingdom of Morocco. It was signed into effect on September 17, 2015. The U.S. Government will provide \$450 million over five years to implement the outlined programs. The MCC is a U.S. foreign aid agency that aims to reduce poverty through investing in development projects in partnership with foreign governments, using a free-market approach. Article I of the Compact states its overarching goal: to reduce poverty in Morocco through economic growth. This objective consists of two projects: the Employability Project, which aims to increase access to workforce training and secondary education, and the Land Project, which seeks to increase land productivity by reducing barriers to investment through formal land markets.³³ The Land Project is further divided into Rural Land Activity (on which this article focuses) and Industrial Land Activity.

In the case of rural land, the Compact focuses on *melkization* of collective land,

which is the process of converting *soulaliyate* land into *melk*, or private property. The Rural Land Activity will function as a pilot project to privatize 46,000 hectares of collective land in the ORMVA Gharb region of Morocco, in order to assess the effectiveness of registration strategies and their ability to be scaled up throughout all irrigated areas in Morocco. In addition to seeking to address productivity constraints, the Compact also identifies governance structures as forming a significant obstacle to investment and productivity increases in collective land. The Compact blames "the lack of a national land governance strategy, policy or sector-wide plan, poor coordination among government actors, and a legal-regulatory framework that is outdated or applied inconsistently" for creating barriers to productive investment in rural land.³³

When addressing productivity issues in *soulaliyate* land, the Compact primarily focuses on tenure insecurity and its associated impact on investment, issues of co-ownership and fragmentation, difficulties in registering and titling collective land under the current system, and transaction obstacles in land markets. The Rural Land Activity seeks to rectify these problems through a new system of registering and titling *soulaliyate* land that will be developed over the first two years of the Compact, as the current system has proven inadequate due to high costs, administrative complexity, and legal barriers to registering collective land.³³ The Compact stipulates that fees and taxes related to registering land must not be an obstacle to the implementation of the outlined projects; the King's guidelines have stipulated that the process must be free for *soulaliyate* members. Following the two-year project development process, this new registration and titling program will be implemented during the latter three years of the MCC grant.²⁰

Although the new process of registration has not yet been fully developed, some details have been released by the Moroccan government. To alleviate potential economic losses to individuals who could lose access to land as a result of the registration process, officials are developing a livelihood restitution plan that will compensate those negatively impacted by this project. It will be based upon socio-economic surveys conducted prior to registration and delimitation.²⁰ Although co-ownership of plots is recognized as a significant obstacle to productive investment and interaction with conventional land markets, it will continue to be allowed following the registration process at the request of rights-holders. Past efforts to combat joint ownership have proven ineffective in Morocco, but the Ministry of Agriculture is in the process of developing

a new approach that will compensate those who exit co-ownership.²⁰

The Compact will be implemented by the MCA-Morocco II, an independent agency of the Moroccan Government that will be led by a Board of Directors with representatives from both state ministries and civil society.³³ The MCC's review process estimates that the planned privatization of land in the Gharb region will affect 81,500 people, who are the rights-holders or family members of those who have rights to collective land in the region.³³ The Government of Morocco and the MCC recognize that the social and economic transformations that will accompany this shift in land tenure may negatively impact historically marginalized groups, especially women. As a result, a section was included in the Compact to signal their commitment to an inclusive reform process through engagement with civil society organizations and potential projects such as legal literacy assistance and inclusive outreach programs.²⁰ However, like many of the potential outlined projects in the Rural Land Activity, a lack of accompanying details or a firm commitment to implementing specific, targeted programs makes it difficult to assess whether such programs will, in fact, occur, and the extent to which they will have a distributional impact on the tenure reform process.

The implementation of this Compact will change the tenure system of collective land in the Gharb region, and this process will eventually be scaled up to cover the 320,000 hectares of identified irrigated collective land in Morocco.²⁰ In addition, the text of the Compact notes that this registration process could be scaled up to cover the approximately 2 million hectares of arable collective land in Morocco, once legal barriers to titling collective land have been addressed.³³ This will have a significant impact on the livelihoods of rural residents of collective areas, which the government estimates at approximately 10 million people, nearly one-third of Morocco's total population.²³

Proponents of privatization and theoretical critiques of market-oriented reform

Systems of land ownership, tenure structures, and related reform processes play a powerful role in shaping the distribution of political, economic, and social power, and are in turn shaped by existing power dynamics and hierarchies. As a result, agrarian systems are a vital area of inquiry for analyzing power structures and distributional relations in rural areas. The debate over agrarian reform was revitalized in the new millennium with the publication of Hernando de Soto's book, *The Mystery of Capital*, which broadly claims

that formalization of property rights and a subsequent creation of national property registers will reduce poverty by allowing informal assets owned by the rural poor to be brought into formal markets.³⁴ Mainstream economists, development policy makers, and international financial organizations are generally in favor of the formalization argument, while heterodox economists and interdisciplinary scholars are critical of it.³⁵

Scholars who subscribe to the formalization model believe that it increases tenure security and prevents the need for property owners to spend scarce resources on unproductive investments to protect their property.³⁶ Furthermore, they argue that tenure security invariably leads to greater productive investment as a result of increased credit access and increases market efficiency by lowering market transaction costs. Finally, and perhaps most significantly, these scholars have claimed that formalization of property rights will directly lead to a more equitable distribution of land, while also allocating land to the most productive users, resulting in economic growth.^{37,38} This argument is predicated on the idea that small, owner-operated farms are more efficient and have better economic outcomes than large-scale agriculture in development projects.³⁹ The policies and projects outlined in the Compact broadly fall within this framework, and much of the Moroccan government's development rhetoric aligns with ideas of economic liberalization.³

Critics of this approach have emphasized that proponents of formalization view land primarily as an economic resource and tend to ignore non-economic dimensions of land, such as identity construction, social relations, and political power.^{40,41} Social factors can mitigate market-based land sales due to community solidarity and identification with a geographic location, but processes of agrarian transformation have the potential to either strengthen or erode community bonds.¹³ Another important critique is that titling programs produce asymmetrical benefits. Actors with economic capital, political connections, and more experience working with national institutions have disproportionately benefited from titling programs.⁴² Skeptics point to a process of "elite capture" in which large landowners and agrarian capitalists use their influence to shape reform programs for their own economic benefit.^{43,44}

Agrarian structures can be characterized by modes of production, social class relations, and land tenure institutions. Broadly, this categorization process divides agrarian structures into capitalist, semi-feudal, communal, and socialist systems.¹⁶ In Morocco, the primary agrarian systems are

communal and capitalist, and government policies and reforms have strengthened the capitalist system at the expense of the communal system by shifting the preferred mode of production and tenure structures.^{2,6} The growth of capitalist structures in the agrarian countryside contributes to the proletarianization of the rural population as wage labor becomes an increasingly important economic activity due to the dominance of the capitalized farming sector.⁶ By converting collective land into private property, the Compact will open up collective land in the Gharb region to market processes and shift modes of production, which will strengthen capitalist structures and stimulate processes of proletarianization.^{9,16}

Impact of privatization and titling on tenure security, investment, and credit access

The idea of formalizing customary tenure has shaped the development of Morocco's privatization project, as lists of customary rights-holders will be compiled to decide who will receive titles to collective land in the Gharb irrigated perimeter.²⁰ However, this view is largely predicated on the assumption that tenure insecurity results purely from lack of formal recognition of private property rights. It is important to note that tenure insecurity can arise from a variety of sources, and policy solutions must be contextually tailored to address the relevant causes. Proponents of market-oriented reform processes have largely ignored these issues.⁴²

The assumed economic benefits of the Rural Land Activity are based upon the idea that tenure security and formal property rights incentivize productive investment in land and facilitate access to credit that can be used for investment by allowing land to be used as collateral.³³ There is a significant body of work that demonstrates the link between tenure security and productive investment, although the results of specific projects have varied based on context.^{30,35,36,37} However, there is little evidence that titling projects increase access to credit for small landowners.⁴⁵ Studies have shown that medium and large landowners experience increased access to credit as a result of receiving land titles.⁴⁶ The literature presents three compelling explanations for this result. First, commercial banks will often only give preferential loans that use plots over a certain size as collateral, which marginalizes smallholders.⁴¹ Secondly, social norms can sanction those who use family or community land as collateral, as land can serve as an important marker of community identity.⁴⁰ Finally, microfinance institutions and informal loans can fill the credit gap for small landowners, which provides a

disincentive to attempt to break into a formal credit market that has historically marginalized small farmers.⁴¹

The Rural Land Activity is working under the assumption that receiving a title provides greater access to credit, but the MCA-II recognizes that structural support will help beneficiaries break into formal credit markets. Farmers in irrigated collective areas have been able to acquire loans, but the terms have typically been prohibitive due to lack of usable collateral.²³ To provide structural support, the MCA-II is developing a partnership agreement with *Crédit Agricole du Maroc* (CAM), a bank that focuses on loans for agricultural development. This partnership will include mobilizing CAM agencies to provide funding for investment projects for the beneficiaries of the Rural Land Activity, grant technical support for farmers, and conduct outreach campaigns to ensure that farmers know about their credit options.²⁰ However, it is important to recognize that there may be a significant gap between the rhetoric of these projects and their outcomes. Agrarian capitalists have long been able to get preferential access to credit, and the government has promoted agricultural policies that subsidize loans for their operations.¹³ This aligns with the regime's particular interpretation of neoliberalism that provides benefits to elites, and a lack of accountability will likely result in greater benefits for agrarian capitalists than small farmers in the ORMVA Gharb.

Legal and administrative implications of privatization

The privatization project outlined in the Compact will change the legal systems and administrative structures regulating collective land in the ORMVA Gharb irrigated perimeter. The shift from management by customary law to statutory law will alter administrative structures, with consequences for both local and national actors and institutions. It is important to detail how these shifts affect different actors involved in the management of collective land, as this will impact power dynamics and hierarchies that have a mutually constitutive arrangement with systems of land management.⁴⁷

Collective land is subject to legal pluralism – in terms of statutory law, it is regulated by the *Dahir* of 1919 and various attached reforms, most notably the legal change in 1969 regarding irrigated areas, as detailed earlier. This statutory law gives ethnic collectivities the right to manage their collective lands by customary law as defined by local councils and *nouab*, under the oversight of the Ministry of the Interior.⁹ The lack of codification of customary law has allowed it to be re-defined by local councils and *nouab*, who have historically re-allocated land to households based on the needs of their

members and ability to work given plots of land.¹⁴ While customary law continues to be dynamic, agricultural *soulaliyate* land is no longer re-allocated due to land scarcity and demarcation of household plots through formal and informal delimitation processes.¹⁷

The proposed *melkization* and titling process will shift administrative oversight of *soulaliyate* land in the Gharb irrigated perimeter from the Ministry of the Interior to the ORMVA Gharb, as part of the Ministry of Agriculture.² According to one government official, the Ministry of the Interior welcomes this loss of management, as the drawbacks of administering collective land have far outweighed any benefits the Ministry has experienced.²³ Furthermore, once collective land formally becomes *melk*, it will no longer be regulated by customary law, and local councils will lose their ability to manage the land of their ethnic collectivity. Nevertheless, the councils will remain powerful actors throughout the implementation process, as they will have a significant role in deciding the criteria for rights-holders and establishing formal lists of beneficiaries to collective land in conjunction with the MCA-II.²⁰

In light of the literature on privatization projects in general, as well as studies of particular cases in Morocco, there are two significant outcomes that appear likely to occur as a result of legal and administrative changes. The first of these is cementing of ownership structures, as connected to the formalization process inherent in titling programs.⁴⁸ While some of the collective land within irrigated areas has been delimited and subdivided through formal and informal means, this has not encompassed all such land within the ORMVA perimeters.^{17,20} Through *remembrement* and allocative processes, the Compact will erase the historically dynamic nature of customary allocative mechanisms and could comprehensively exclude those who had informal access to land as the system shifts to an explicit private property regime.⁴⁹

A second notable outcome will be to reduce legal and administrative barriers to transferring land in irrigated areas through conventional land markets.^{20,33} The conversion of collective land into *melk* removes legal barriers that in the past have prevented *soulaliyate* land from interacting with conventional land markets. There are contentious debates among scholars of land privatization about the outcomes of market-based allocation of land; various case studies have shown differing impacts as a result of asymmetric reform processes, agrarian structures, and the relative power of different vested interests.^{5,36,41,42,45}

Market-based allocation and land concentration

The Compact explicitly states that “the Land Strategy will be market-oriented.”²³³ This comes as no surprise, as it is well-recognized that the policies of major development organizations have largely worked under a liberal economic framework that promotes market processes.^{50,51} With the conversion of collective land into *melk*, the beneficiaries of the project will be able to sell and lease the land they have a title to within the irrigated perimeter through formal market systems. While some constraints will remain due to ORMVA policies, the most significant legal barriers to transferring land through market systems will be alleviated through the privatization process.²⁰

Proponents of marketization argue that markets are the most efficient institution to allocate land to the most productive users with the fewest transaction costs, in the absence of market imperfections such as information asymmetries, government-mandated land ceilings, and insecure property rights.^{37,44} However, even scholars working under this framework acknowledge that in many areas of the developing world where market-oriented reform projects are carried out, there is a lack of good governance, institutional capacity to protect property rights, or systems in place that prevent information asymmetries.³⁶ Nevertheless, those who work under this framework still believe that the market is the best institution to allocate land and reforms can be designed to fix market imperfections.⁴³

There are a number of critiques of market-led reform projects that are relevant to the Compact’s proposed privatization scheme. First, critics accuse proponents of marketization of ignoring historical power imbalances that have shaped the development of contemporary land ownership structures.⁵ Markets are embedded institutions constructed through social and political processes and embody societal power hierarchies.⁵² Holders of liquid capital have increased power in market systems, as they have the disposable income to make productive investments.⁴⁰ We can certainly see this in the development of an agrarian capitalist class during the colonial period; their wealth makes them the primary group with the capital to invest in technology, agricultural inputs, and land, which gives them more power in land markets than small farmers.^{6,13}

A second critique of market-led land reform is its potential to concentrate land ownership.^{41,43} This is a concern for both proponents and critics of liberalization and market processes, as there is evidence that small farmers are the most productive users of agricultural land.³⁹ Land concentration

has been a continuing characteristic in Morocco’s land system since the colonial era; at the time of independence, some 500,000 rural Moroccan households were virtually landless.¹⁹ Furthermore, approximately 500,000 small farmers disappeared between the General Agricultural Census of 1974 and 1996.¹ With regard to irrigated land, 4% of farmers own half of all irrigated land in Morocco.⁶ One report noted that between 1965 and 1970, some 12,000 peasants became landless in the Gharb valley while 100 new capitalist landowners appeared.⁶ This process is driven by a combination of population growth, land scarcity, rural out-migration, and agricultural policies that promote consolidation and prioritize large land-holdings.^{2,9,50}

Land concentration and consolidation are also tied to distress sales. Critics of formalizing property rights note that removing barriers to land markets allows those in economic distress to sell their land in order to quickly raise funds.⁴¹ This has historically been most notably seen in Morocco during times of drought, when small-farmer agriculture becomes untenable as a livelihood strategy.² In addition, the Gharb region could be particularly susceptible to distress sales, as the poverty level in the region was 20.5% in 2008, compared to the national average of 14%.¹³ However, the corollary argument is that selling part or all of a plot of land can raise capital to invest in alternative income-generating activities. Some rights-holders in the Gharb region report that farming alone cannot support them, and that they would use capital to invest in other livelihood strategies.¹³

In formal market systems, only large- or medium-sized landowners have the capital to invest in acquiring more land once legal barriers to market transactions are removed.⁴³ This is further compounded by the greater ability of large landowners to access credit, as detailed above. Land concentration will enhance the relative stature of the modern agricultural sector, which will in turn strengthen agrarian capitalist structures in the rural countryside.⁹ At the same time, this will increase the market for agricultural wage labor, as the agrarian capitalist class are largely absentee landowners and use wage labor to work their operations.¹³ The potential increase in the rural landless population would coincide with an increased demand for agricultural labor, allowing for livelihood transformation.^{6,7} In this way, the growth of capitalized agriculture will create space to incorporate some of those displaced by its very expansion.

It is important to recognize that this is not a monolithic process, nor is it a simple matter of an agrarian capitalist class

gaining economic power at the expense of rural farmers in the traditional sector. Small-holder farming continues to play an important role in many regions of Morocco, especially outside of irrigated perimeters where collective land is still managed by the *Dahir* of 1919. Approximately half of Morocco's population is employed, formally or informally, in the agricultural sector.²⁴ Furthermore, there is increasing evidence that small-holders are diversifying their livelihood strategies. In pastoral areas in eastern Morocco, sedentarization is occurring and members of ethnic collectivities are claiming land for agriculture, but this process is not an elite phenomenon nor has it eliminated the importance of the livestock economy.¹⁸ In the southeastern steppes, research shows that land-use change related to commercial agriculture is integrated with non-capitalist processes and that households are increasingly diversifying their livelihood strategies while maintaining historical practices of peasant farming.^{4,31}

Political considerations and elite capture of reform benefits

The current Moroccan regime has had a mutually beneficial relationship with landowning rural elites, who have supported the monarchy in exchange for patronage.³ A common theme in the literature on tenure reform is the role that elites have played in shaping institutional change for their own benefit, as well as mitigating reform processes that would weaken their power.^{5,35,44} This has certainly played out in Morocco in past privatization schemes and will likely continue to shape reform processes throughout the Compact's implementation.⁵³

The Moroccan government undertook some land reform in the late 1950s and early 1960s, when it nationalized colonized land.² However, much of the land that was supposed to be nationalized was illegally sold by former colonists to Moroccan elites.⁶ Nearly two-thirds of all former colonized land went to these elites: 410,000 hectares were sold directly by French owners, with another 163,000 hectares sold to elites through a privatization scheme in the 1980s.² The concession of nationalized land to private interests can be interpreted as a continuation of the systems of patronage that date back to pre-Protectorate political regimes.¹ It has contributed to the development of an agrarian capitalist system and co-opted agrarian elites for political support.⁹

Land speculation is one area where political connections have assisted Moroccan elites in capturing economic benefits from irrigation schemes. Given that land improved by irrigation rises enormously in value, those who purchase land within an irrigation

perimeter prior to the project announcement or implementation can gain significantly on their initial investment. In certain cases, advance information about future irrigation projects was allegedly leaked to some individuals; in the Souss region, three individuals purchased 40% of land that was to be irrigated by the Massa Dam prior to its construction.^{2,54} Political connections can help elites take advantage of infrastructure projects as well as reform processes for their own economic benefit through access to information that is not publicly available.³⁵

While it is difficult to identify the extent to which the 2011-2012 protests contributed to or played into the timing of the Compact, it is clear that the regime has deployed rhetorical strategies to emphasize that the projects outlined in the Compact will provide economic benefits for small farmers in the Gharb irrigated perimeter.²³ In this way, the privatization of collective land will serve the interests of the regime in placating both agrarian capitalists and small farmers. All of the *soulatiyates* whom I interviewed were in support of privatization, although there were concerns about how it would be implemented. These concerns were primarily raised by widows and divorced women, who stood to comprehensively lose access to land as a result of privatization that aligned with customary law.

The specter of agrarian reform remains a potent political tool for the Moroccan state, both rhetorically for public consumption as well as practically within the elite political sphere. For the rural poor and the political left, reform serves as a promise for greater equity; for the regime, it is a useful threat against the landholding elite in order to ensure their loyalty.² As a whole, this has worked out to the benefit of agrarian elites at both local and national levels, who have been able to leverage the regime's need for support in order to gain economic benefits, particularly in the accumulation of large landholdings and the development of policies aimed at promoting their large-scale capitalized agricultural operations.^{1,13}

Conclusion

The Rural Land Activity included in Morocco's Employability and Land Compact with the MCC will only directly impact a relatively small area of collective land in Morocco through a privatization process that will give titles to collective rights-holders. However, it can be placed within a longer history of agricultural policies that have been structured according to neoliberal ideas that emphasize liberalization and marketization and focus on issues of modernization, economic productivity, and allocative efficiency. Historically, the benefits of these policies have accrued to agrarian capitalists

who dominate the modern agricultural sector, while the traditional sector has largely been ignored or subsumed by these policies. Although many of the specific details of the projects outlined in the Compact are still being designed by the Government of Morocco in conjunction with the MCC, we can situate some potential outcomes of the proposed privatization of collective land in the Gharb irrigated perimeter within larger processes of agrarian transformation that are occurring in the Moroccan countryside.

First, the modern agricultural sector will continue to grow, as ORMVA institutions will focus on modernizing farms on collective land as part of the privatization process. This will include technical training and support from the Office National du Conseil Agricole through a partnership agreement with the MCA-II, encouraging *soulatiyate* farmers to shift from cereal production to fruit plantations and organizing farmers to negotiate with large-scale distributors, further connecting them with capitalized networks.²⁰ This is aimed at increasing the productivity of these plots and supporting productive investment in order to increase farmer incomes. However, the shift from cereals to fruit cultivation will further decrease the availability of subsistence agricultural goods in a country that is already food insecure.¹¹

Second, the removal of legal barriers to marketization of collective land in the Gharb region will allow for land concentration, though it does not make it inevitable. The privatization and titling process will greatly increase the value of collective land within the irrigated perimeter, which could encourage small farmers to sell or rent their land to larger operations that are run by agrarian capitalists, which has already been documented as a trend in the region.¹³ The extent to which land consolidation and concentration occurs within the Gharb in the post-reform process will be an important area of research in order to assess how this privatization process has shifted control over the land, and whether it will continue to be farmed and owned by the same population.

Finally, as the modern sector increases in prominence, small-holder subsistence farming is becoming increasingly untenable as the sole livelihood strategy for the rural population.⁵ While this does not necessarily mean that small rural landowners will sell their land and move to cities or completely shift to wage labor, we can see that livelihood transformations are occurring in various rural areas in Morocco. The greater importance of wage labor, particularly in agriculture, will continue to be seen as large operations expand and seek to increase their labor pool. Small rental arrangements, remittances, and mixed pastoral and agricultural livelihood

strategies are playing an increasingly important role in the livelihoods of the rural population, although small farming continues to be a significant occupation.¹³

These dynamics are playing out asymmetrically across Morocco, but shifts in land usage can be seen in many areas, particularly as population growth leads to increasing land scarcity. Socio-economic surveys and follow-up field research will help assess the extent to which this privatization project increases the role of agrarian capitalist systems in formerly collective land within irrigated perimeters by concentrating land ownership and stimulating a diversification of livelihood strategies. The history of agricultural policies favoring large operations makes it clear that capitalist structures are continuing to gain influence in Morocco's agrarian system, and that shifting systems of land tenure play a powerful role in shaping ownership structures and the tenable livelihood options of the rural population. This will have a profound effect on the future of rural Morocco, as it appears that many of benefits of agrarian transformation are accruing to agrarian capitalists, at the expense of small rural farmers.

Acknowledgements

I would like to thank all of the research participants and interviewees who were willing to speak with me in Morocco, without whom this project would not have been realized. Brad Dillman, Nick Kontogeorgopoulos, and Emelie Peine provided me with invaluable advice and comments throughout the research and writing process. I am grateful for the generous funding for fieldwork provided by the International Political Economy Department at the University of Puget Sound. Finally, I am thankful for the hospitality of the Moroccans who welcomed me into their homes and provided logistical support during the fieldwork process.

References

¹Mahdi, M. (2014) "The Future of Land Tenure in Morocco: A Land Grabbing Case." *New Medit.* 13. 4. Pg 2-10.
²Swearingen, W. (1987) *Moroccan Mirages: Agrarian Dreams and Deceptions, 1912-1986*. Princeton University Press.
³Desrués, T. (2005) "Governability and Agricultural Policy in Morocco: Functionality and Limitations of the Reform Discourse." *Mediterranean Politics.* 10. 1. Pg 39-63.
⁴Rignall, K. and Kusunose, Y. (2016) "Governing Livelihood and Land Use Transitions: The Role of Customary Tenure in Southeastern Morocco." Paper presented at the 17th World Bank Conference on Land and Poverty, Washington, D.C., March 14-18.
⁵Bush, R. (2007) "Politics, Power and Poverty: Twenty Years of Agricultural Reform and Market Liberalisation in Egypt." *Third World Quarterly.* 28. 8. Pg 1599-1615.
⁶Daoud, Z. (1981) "Agrarian Capitalism and the

Moroccan Crisis." *MERIP Reports.* 99. Pg 27-33.
⁷Springborg, R. (1990) "Agrarian Bourgeoisie, Semiproletarians, and the Egyptian State: Lessons for Liberalization." *International Journal of Middle East Studies.* 22. Pg 447-472.
⁸Swearingen, W. (1992) "Drought Hazard in Morocco." *Geographical Review.* 82. 4. Pg 401-412.
⁹Bouderbala, N. (1999) "Les systèmes de propriété foncière au Maghreb. Le cas du Maroc." In *Politiques foncières et aménagement des structures agricoles dans les pays méditerranéens: à la mémoire de Pierre Coulomb*, ed. Jouve, A. and Bouderbala, N. Montpellier. Pg 47-66.
¹⁰Zartman, W. (1963) "Farming and Land Ownership in Morocco." *Land Economics.* 39. 2. Pg 187-198.
¹¹El-Ghonemy, M. (1993) "Food Security and Rural Development in North Africa." *Middle Eastern Studies.* 29. 3. Pg 445-466.
¹²Venema, B. (2002) "The Vitality of Local Political Institutions in the Middle Atlas, Morocco." *Ethnology.* 41. 2. Pg 103-117.
¹³Minoia, P. (2012) "Mega-Irrigation and Neoliberalism in Postcolonial States: Evolution and Crisis in the Gharb Plain, Morocco." *Geografiska Annaler: Series B, Human Geography.* 94. 3. Pg 269-286.
¹⁴Michel, N. (2002) "The Individual and the Collectivity in the Agricultural Economy of Pre-colonial Morocco." In *Money, Land and Trade: An Economic History of the Muslim Mediterranean*, ed. Hanna, N. London. Pg 15-36.
¹⁵Wellhofer, S. (1990) "Agrarian Capitalism, Property Rights, and Rural Class Behavior." *Comparative Political Studies.* 22. 4. Pg 355-396.
¹⁶De Janvry, A. (1981) "The Role of Land Reform in Economic Development: Policies and Politics." *American Journal of Agricultural Economics.* 63. 2. Pg 384-392.
¹⁷Jkaoua, A. (2011) "Management of the Collective Lands in Morocco." Presentation at FIG Working Week, Marrakech, Morocco, May 18-22.
¹⁸Kreuer, D. (2011) "Land Use Negotiation in Eastern Morocco." *Nomadic Peoples.* 15. 1. Pg 54-69.
¹⁹Bouderbala, N., Chraïbi, M. and Pascon, P. (1974) "La Question Agraire au Maroc." *Bulletin Economique et Social Du Maroc.* nos. 123-125. Pg 5-423.
²⁰El Ghrib, A. (2016) (a) "Agricultural Collective Land Privatization in Morocco." Paper presented at the 17th World Bank Conference on Land and Poverty, Washington, D.C., March 14-18.
²¹World Bank. (1966) *The Economic Development of Morocco*. Baltimore: John Hopkins University Press.
²²Delanoy, H. (1969) "Le Code des Investissements Agricoles: Nouveau Cadre Législatif de la Transformation des Campagnes Marocaines." *Revue de Géographie du Maroc.* 16. Pg 151-157.
²³El Ghrib, A. (2016) (b) "Agricultural Collective Land Privatization in Morocco." Session at the 17th World Bank Conference on Land and Poverty, Washington, D.C., March 14-18.
²⁴USAID. (2011) USAID Country Profile: Property Rights and Resource Governance, Morocco.
²⁵Rignall, K. (2012) "Land, Rights, and the Practice of Making a Living in Pre-Saharan Morocco." University of Kentucky.
²⁶Recensement Général de l'Agriculture (RGA). (1996) "The Future of Land Tenure in Morocco: a Land Grabbing Case." *New Medit.* 13. 4. Cited in Mahdi M. above (2014). Pg 2-10.
²⁷Personal Interviews with ANCFCC officials, June 2015, Rabat, Morocco.
²⁸Wheeler, S. and Balgley, D. (2016) "The Gendered Implications of Land Tenure Reform in Morocco." Paper presented at the 17th World Bank Conference on Land and Poverty, Washington, D.C., March 14-18.
²⁹Personal Interviews with DAR officials, June 2015, Rabat, Morocco.
³⁰M'Hassni, M., Felji, M. and Khalali, H. (2003) "Le système foncier au Maroc: Une sécurité et un facteur de développement durable. Au milieu urbain et rural." Paper presented at the 2nd FIG Regional Conference,

Marrakech, Morocco, December 2-5.
³¹Rignall, K. (2015) "Land and the Politics of Custom in a Moroccan Oasis Town." *Anthropological Quarterly.* 88. 4. Pg 941-968.
³²Scott, J. (1998) *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Binghamton, NY: Vail-Ballou Press.
³³Millennium Challenge Corporation. (2015) *Morocco Employability and Land Compact*. Washington, DC.
³⁴De Soto, H. (2000) *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else*. New York, NY: Basic Books.
³⁵Sjaastad, E. and Cousins, B. (2008) "Formalisation of Land Rights in the South: An Overview." *Land Use Policy.* 28. Pg 1-9.
³⁶Deininger, K. and Feder, G. (2009) "Land Registration, Governance, and Development." *World Bank Research Observer.* 24. 2. Pg 233-266.
³⁷Deininger, K. and Binswanger, H. (1999) "The Evolution of the World Bank's Land Policy: Principles, Experience, and Future Challenges." *The World Bank Research Observer.* 14. 2. Pg 247-276.
³⁸Levine, R. (2005) "Law, Endowments and Property Rights." *Journal of Economic Perspectives.* 19. 3. Pg 61-88.
³⁹Manji, A. (2003) "Capital, Labour and Land Relations in Africa: A Gender Analysis of the World Bank's Policy Research Report on Land Institutions and Land Policy." *Third World Quarterly.* 24. 1. Pg 97-114.
⁴⁰Akram-Lodhi, A. (2007) "Land, Markets, and Neoliberal Enclosure: An Agrarian Political Economy Perspective." *Third World Quarterly.* 28. 8. Pg 1437-1456.
⁴¹Musembi, C. (2007) "De Soto and Land Relations in Rural Africa: Breathing Life into Dead Theories about Property Rights." *Third World Quarterly.* 28. 8. Pg 1457-1478.
⁴²Otto, J. (2009) "Rule of Law Promotion, Land Tenure and Poverty Alleviation: Questioning the Assumptions of Hernando de Soto." *Hague Journal on the Rule of Law.* 1. Pg 173-195.
⁴³Lahiff, E., Borras, S. and Kay, C. (2007) "Market-Led Agrarian Reform: Policies, Performance and Prospects." *Third World Quarterly.* 28. 8. Pg 1417-1436.
⁴⁴Loehr, D. (2012) "Land Reforms and the Tragedy of the Anticommons—A Case Study from Cambodia." *Sustainability.* 4. 4. Pg 773-793.
⁴⁵Deininger, K. (2010) "Towards Sustainable Systems of Land Administration: Recent Evidence and Challenges for Africa." *African Journal of Agricultural and Resource Economics.* 5. 1. Pg 205-226.
⁴⁶Williamson, C. (2011) "The Two Sides of de Soto: Property Rights, Land Titling, and Development." *The Annual Proceedings of the Wealth and Well-Being of Nations.* Pg 95-108.
⁴⁷Brara, J. (1983) *The Political Economy of Rural Development*. New Delhi: Allied Publishers Private Limited.
⁴⁸Deere, C. & León, M. (2001). Who Own the Land? Gender and Land-Titling Programmes in Latin America. *Journal of Agrarian Change*, 1(3), 440-467.
⁴⁹Demsetz, H. (1967) "Toward a Theory of Property Rights." *The American Economic Review.* 57. 2. Pg 347-359.
⁵⁰Davis, D. (2006) "Neoliberalism, Environmentalism, and Agricultural Restructuring in Morocco." *The Geographical Journal.* 172. 2. Pg 88-105.
⁵¹Doss, C., Summerfield, G. and Tsikata, D. (2014) "Land, Gender, and Food Security." *Feminist Economics.* 20. 1. Pg 1-23.
⁵²Razavi, S. (2007) "Liberalisation and the Debates on Women's Access to Land." *Third World Quarterly.* 28. 8. Pg 1479-1500.
⁵³Najem, T. (2001) "Privatization and the State in Morocco: Nominal Objectives and Problematic Realities." *Mediterranean Politics.* 6. 2. Pg 51-67.
⁵⁴Oussalah, L. (1976) "Le Développement Inégal au Niveau National: Maroc utile/Maroc inutile." *Mémoire*. Aix-en-Provence: University of Aix-Marseille II.

Work-Related Soft Skills Summer Program: A Case Study

BY J. CHRISTIAN BANEZ & SHAYDA AFRASSIAB

TRUMAN STATE UNIVERSITY

MENTORS:

CAROL COX, PROFESSOR/HEALTH SCIENCE

TRUMAN STATE UNIVERSITY

JOSEPH VISKER, ASSISTANT PROFESSOR/HEALTH SCIENCES

MINNESOTA STATE UNIVERSITY-MANKATO

Abstract

Employment issues such as finding full-time jobs with wages that allow for independent living are often worse for youth with disabilities. Lack of soft or social skills has been noted as a major barrier to their workplace success. For youth with cognitive disabilities, not having soft skills places them at a disadvantage in transitioning from school to work. This exploratory study describes a trial summer-long program that focuses on improving school-to-work transition difficulties of youth with cognitive disabilities. This program included three phases: a paid work experience for four days each week, a classroom-based soft skills curriculum for one day each week, and a culminating school-to-work transition training event at the end of the summer. The program followed best practice recommendations. A significant ($p < .05$) positive change in participant self-reported acquisition of soft skills was noted pre-post-program. Comprehensive programs that follow best practice recommendations and include transition support services may be effective in soft skills improvement and may assist in overcoming an employment barrier for this population.

Introduction

Youth with cognitive disabilities are legally entitled to transition services. Employment issues, including school-to-work transition, are often worse for youth with disabilities. After high school, most still live with their parents and are either underemployed or unemployed.^{1,2} When experts, parents, and youth with disabilities were interviewed about important skills for youth with disabilities to learn, employment skill-building for school-to-work transition was rated as an important employment-directed resource.³

People with cognitive disabilities must have the same opportunity as other citizens to become active and valued community members.⁴ As they transition from school to work, many youth with cognitive or intellectual disabilities lack access to opportunities and experiences that would allow them to advocate for themselves in the area of employment and financial decisions.⁵ Data from the National Longitudinal Transition Study-2 showed that many youth with cognitive disabilities were not very likely to take control of their own school-to-work transition planning.² In general, support agencies also assumed limited roles in transition planning and youth with cognitive disabilities needed more support services

than those with other disabilities. Although half seem to find employment after high school, only about 70% were employed in full-time positions.² When successful support services and interventions for youth with disabilities were implemented, future adult employment levels and community resource access were improved.⁶

Successful employment for those with disabilities is associated with employer perception towards inclusion.⁷ Lack of employment skills, including soft (e.g., social, interpersonal, communication) skills have been noted as major barriers to their workplace success.^{1,8} Soft skills are viewed as increasingly valuable by employers of people with disabilities.⁹ Vocational counselors and potential employers for entry-level positions for teens were surveyed and noted positive attitude, flexibility, and soft skills as important for transitioning from school to work.¹⁰ In another study, for those participating in vocational rehabilitation, the most important work-related soft skills included body language and socializing with others.¹¹ It seems, too, that soft skills were a relatively strong predictor for post-high school employment for youth with disabilities.¹²

Best practices in school-to-work transition vocational rehabilitation services for youth with cognitive disabilities

include employment with ongoing support services. In addition, secondary supports, like transition services, as well as soft skills training, have gained promise. It has been recommended, however, that the effects of soft skills training and ongoing support be more thoroughly assessed.¹³ For example, in one state, participating in vocational rehabilitation services by youth with cognitive disabilities seemed to positively influence employment.¹¹ In another state, participating in more customized vocational rehabilitation with contracted employment demonstrated increased effectiveness in moving youth with disabilities toward employment as compared to traditional vocational rehabilitation.¹³

Most vocational rehabilitation agencies use personal social skills training and environmental supports as interventions to improve soft skills in their consumers.⁸ Social skills training, role playing social situations and receiving feedback from instructors; is commonly used to develop relationship and problem-solving skills in youth, including those with disabilities.¹⁴ In a meta-analysis, this type of social skills training for youth with cognitive disabilities demonstrated moderate effects, and the school setting seemed to be better able to meet youth needs than the workplace setting.¹⁵ The purpose of this case study was to assess change

in participant self-reported acquisition of soft skills after a trial summer-long program that focused on improving school-to-work transition difficulties of youth with cognitive disabilities.

An agency that provides employment services coordination and community integration for people with disabilities offered an eight-week program during the summer using paid work experience, a day-long school-to-work transition training event, and a standardized work-related soft skills curriculum for adolescent youth and young adults with cognitive disabilities. This program focused on supervised work experience and work-related soft skills, as well as instruction in self-knowledge, independent living, working, and becoming a greater part of their community. This case study reports on the three program phases (e.g., paid work experience, transition planning training event, work-related soft skills curriculum) and participants' change in self-reported acquisition of soft skills.

Method

Participants

Thirty youth with cognitive disabilities attended an eight-week, paid, summer work experience including standardized work-related soft skills curriculum and a transition training event hosted by a regional support agency for people with disabilities. The agency advertised this summer program on their website and encouraged consumers who met the criteria to attend. In order to qualify for the program, participants possessed low-moderate cognitive functioning with deficits in one or more adaptive behaviors (living/communication/social skills). Participants were between the ages of 16- 21 (28/30 high-school aged), all were white, and 63% (19/30) were male.

Setting

The agency that provided the employment services was headquartered in a rural area of a Midwest state. The agency contracted with six nearby workplaces including school districts and small business establishments to employ the youth participants.

Survey

After Institutional Review Board approval, summer program participants were also invited to participate in a small pre-post program survey. All attendees (with parental/guardian consent and participant assent) agreed to also participate in the survey portion of the program. The Soft Skills Assessment for Secondary Students (SSA) was used by youth participants to self-assess their pre-post program change in work-related soft skills.⁷ Using a modified Likert-type scale, the first question asked

about their self-perceived awareness level of soft skills. The next eight asked about their self-perceived levels of using the soft skills of communication, confidence, and rapport. Possible total scores ranged from 0-32 with higher scores indicating higher perceived levels of using soft skills. Internal consistency reliability (Cronbach's alpha) analyses were conducted for both the pre- and post-tests, yielding scores of $\alpha=0.623$ and $\alpha=0.580$, respectively.

Program phases

Phase 1: paid work experience. A pre-program organizational meeting was held for all small group supervisors who would serve as adult mentors for six groups of five participants as they completed their supervised work experiences and agency staff. Small group supervisors were trained in how to administer the pre-post SSAs in the classroom setting including the assent form. To insure confidentiality, no names were recorded, and code numbers were used on the instruments for all youth participants. After the first week of the program, the small group supervisors administered the Pre-SSA to all youth participants in a classroom or meeting room setting following the specific directions and check sheets from their training. Youth participants were allotted all the time they needed to complete the survey, and the agency staff provided any accommodations. Most participants completed the survey in less than one half hour. The surveys were collected and placed in a sealed envelope and given to a researcher.

To determine the most appropriate workplace for each participant, agency staff, small group supervisors, parent/guardians, and participants met to look through the list of workplaces that volunteered to host participants. Once placements were determined, four days each week over eight weeks, participants in their small groups worked for eight hours each day at one of six work sites (e.g., small businesses or school districts) under the supervision of a site supervisor and the small group supervisor. At the small businesses, participants stocked shelves, cleaned, conducted clerical duties, bagged retail items, and assisted with inventory. At the school districts, they did landscaping, cleaning, maintenance assistance, and working with a team. Also, as a typical employee would, participants were given two 15-minute breaks and a lunch break during their daily shift. Each week, small group supervisors and site supervisors provided oral and written work-related job skills feedback to participants regarding their satisfaction with the participants' workplace skills and behaviors. Check sheets were completed on each

participant by the supervisors who rated them on independence level in performing their clerical, maintenance or other tasks as well as their behaviors such as timeliness, appropriate dress, working with others on a team, and coping with set-backs. Each week, supervisors met with participants and discussed their observations either formally or informally.

Phase 2: standardized work-related soft skills curriculum. One day each week for a half day, all participants in their small groups attended classroom lessons in the Skills to Pay the Bills educational curriculum taught by agency staff.¹⁶ The curriculum, created by the Office of Disability Employment Policy, focused on work-related interpersonal soft skills for teens and young adults, including those with disabilities. Although no staff supervisors observed to ensure fidelity, all staff reported following the lesson plans provided. Instruction included six modules that used active-learning strategies for workforce readiness/soft skills topics including; module 1: communication – providing and receiving information; module 2: attitude- enthusiasm and positive thinking in the workplace; module 3: teamwork – roles and conduct as a team member; module 4: networking – interview guidelines and social networks; module 5: critical thinking, problem-solving, and decision-making; and module 6: professionalism – integration of all of the module skills. To meet the needs of youth with disabilities, the curriculum directed instructional staff to use a variety of teaching methods targeted to all learning styles.¹⁶

Phase 3: day-long, school-to-work transition training event. In addition to the soft skills curriculum, a one-day school-to-work transition planning event was held towards the end of the summer program. Taught by agency staff, the event included sessions, games, and activities. Sessions covered employment preparation, vocational development, self-determination, self-advocacy, independent living, and community engagement using lecture and discussion methods. Session content was then immediately applied through skills practice. For example, participants engaged in communication and job interview role playing activities to apply what they have learned. As a culminating activity, participants created goals for future employment as well as used transition-planning tools such as self-assessments, planning guides, job resource work sheets, job interview tips handouts, transitional living resources list, and lists of support agencies and services in the area to chart a path to meet those goals. On the last day of the program, small group supervisors administered the Post-SSA to all youth participants in a classroom or meeting room

setting following the specific directions and check sheets from training. The surveys were collected and placed in a sealed envelope and given to a researcher.

Results

Paired-samples t-tests were conducted for only those tests that could be matched pre-post, and aggregated means were then compared. Initially, all 30 youth participants completed the pre-SSA; however, only 16 completed the post-SSA. Some did not complete their post-tests due to social-emotional difficulties. Small group supervisors reported that during the post-test session, some youth participants were having behavioral and emotional difficulties that day. Only matched pre-post SSAs, therefore, were assessed. The results of the paired-samples t-test, assessing differences in soft-skills scores between the pre- and post-SSAs ($n=16$), revealed a statistically significant difference between the pre-test ($M=20.56$, $SD=4.98$) and the post-test ($M=23.56$, $SD=4.13$) ($t(15)=-3.105$, $p<.05$).

Discussion

The purpose of this case study was to assess change in participant self-reported acquisition of soft skills after a trial summer-long program that focused on improving school-to-work transition difficulties of youth with cognitive disabilities. Following best practice recommendations for employment support, a summer-long program was conducted and included three phases: a paid work experience for four days each week, a classroom-based soft skills curriculum for one day each week, and a culminating school-to-work transition training event at the end of the summer. A significant positive change in participant self-reported acquisition of soft skills was noted pre-post-program.

Limitations

In regards to the utilization of the SSA and evaluating individual attainment of soft skills, there is always a risk in using a self-report instrument. While there is no specific reason to believe that participants were not truthful about their evaluations of their own abilities, participant self-assessment could have introduced personal bias in their skills. Additionally, this program was conducted in a single location in a single state, thus, compromising the generalizability of the results. While the pre-post assessment design utilized in this study was sufficient to assess changes in participant self-reported soft skills, future programs should include an adequate control group design or supervisor observation to better determine if changes were truly attributed to the program itself.

It is also important that the shortcomings of the instrument used in this study are addressed. Future studies should continue to examine Cronbach's alpha scores in the SSA among different populations to determine whether the internal consistency reliability of the instrument is truly acceptable.

Implications

Lack of soft skills in youth with cognitive disabilities places them at a disadvantage in transitioning from school to work.¹⁰ Participant self-report of soft skills acquisition in this program significantly increased pre- to post-program providing some support for the comprehensive vocational, curricular, and special event style of programming implemented by this agency. Best practices recommend ongoing support and social skills training which this program included with its small group mentors, staff support, planning event, and soft skills curriculum.¹¹ In addition, summer is a good time for those with cognitive disabilities to obtain this extra work experience and continuing support and training.

Most vocational rehabilitation programs for people with disabilities emphasize social skills training and community supports, especially in the school-based setting, with modest results.¹⁵ This program, though, was conducted mainly in the work setting with classroom-style lessons only once each week and a full-day transition planning event. Reinforcement and application of curricular and work-setting lessons was also provided during the special event as participants set goals and planned for their future employment. Providing work and social skills training using a summer program can be attempted by other employment or vocational rehabilitation agencies to continue to improve on any positive results gained during the school year.

The format of the program with its lecture-lab arrangement and the special event that tied all of the pieces together may help to explain the improvement in participant self-reported soft skills. Use of the standardized curriculum with active-learning strategies possibly reinforced the social skills messages received from the small group and employer supervisors every day at work. The opportunity to practice skills in each of the curricular topic areas that were similar to their daily work experiences may have emphasized the importance of workforce readiness and social skills to the participants.

Future research should be done to determine the effects of soft skills training and ongoing support in vocational rehabilitation.¹⁰ When soft skills were assessed in participants pre- to post-program, significant positive change in self-reported acquisition of soft

skills was noted. Conclusions that can be drawn are limited, and caution should be exercised in interpretation of the results, due to small sample size. Many youth with cognitive disabilities, however, lack access to activities and opportunities such as this program.⁴ If results of this exploratory trial program are established in larger programs, comprehensive programs such as this one may assist in overcoming an employment barrier in this population. In conclusion, agencies that provide employment services for people with disabilities do not take large roles in transition planning.² However, in this program, the agency not only provided work placement and a skills-based based curriculum for youth participants, but also planned a full-day event using active-learning methods to teach important transition skills. Youth with cognitive disabilities need more support services than youth with other disabilities.² If results of more complete studies confirm the preliminary results found in this program, it seems that comprehensive programming that includes transition support services may be effective in soft skills improvement. Although it may take more time and effort for agencies to plan and implement more wide-ranging programming, it may better meet the needs of youth and young adults with cognitive disabilities.

References

- ¹Riesen, T., Schultz, J., Morgan, R., and Kupferman, S. (2014). "School-to-Work Barriers as Identified by Special Educators, Vocational Rehabilitation Counselors, and Community Rehabilitation Professionals" *Journal of Rehabilitation* 80.1. Pg. 33-44.
- ²Shogren, K., and Plotner, A. (2012). "Transition Planning for Students with Intellectual Disability, Autism, or other Disabilities: Data from the National Longitudinal Transition Study-2", *Intellectual and Developmental Disabilities* 50.1. Pg.16-30.
- ³Caenepeel-Knust, J. (2013). "Stepping Up: Employment Skills and Support for Students with Disabilities" Project: Master of Arts Degree in Education. California State University San Marcos.
- ⁴American Association on Intellectual and Developmental Disabilities. (2008). "Policy statement: Self-determination." *American Association on Intellectual and Developmental Disabilities (AAIDD)*. American Association on Intellectual and Developmental Disabilities (AAIDD). <https://aaid.org/news-policy/policy/position-statements/self-determination#.VqOzL8uYbes> (Accessed 02/05/2016).
- ⁵Gragoudas, S. (2014) "Preparing Students with Disabilities to Transition from School to Work through Self-Determination Training", *Work and Disability* 48.3. Pg. 407-411.
- ⁶Heera, S. (2016) "Employer's Perspective towards People with Disabilities: A Review of the Literature." *The South East Asian Journal of Management* 10.1 Pg. 54-74.
- ⁷Schochler, S. (2014) "Soft skills assessment for secondary students." *Academia*. Academia. https://www.academia.edu/9614516/Soft_Skills_Assessment_for_Secondary_Students_Introduction (Accessed 02/01/2016)
- ⁸Phillips, B., Kaseroff, A., Fleming, A., and Huck, G. (2014) "Work-related social skills: Definitions and interventions in public vocational rehabilitation." *Rehabilitation Psychology* 59.4. Pg. 386-398.
- ⁹Lindsay, S., Adams, T., Sanford, R., McDougall, C., Kingsnorth, S., and Menna-Dack, D. (2014) "Employers' and Employment Counselors' Perceptions of Desirable Skills for Entry-Level Positions for Adolescents: How Does it Differ for Youth with Disabilities?" *Disability and Society* 29.6. Pg. 953-967.
- ¹⁰Leahy, M. J., Chan, F., Lui, J., Rosenthal, D., Tansey, T., Wehman, P., et al. (2014) "An Analysis of Evidence-Based Best Practices in the Public Vocational Rehabilitation Program: Gaps, Future Directions, and Recommended Steps to Move Forward." *Journal of Vocational Rehabilitation* 41.2. Pg.147-163.
- ¹¹Jun, S., Kortering, L., Osmanir, K., and Zhang, D. (2015) "Vocational Rehabilitation Transition Outcomes: A Look at One State's Evidence." *Journal of Rehabilitation* 81.2. Pg. 47-53.
- ¹²Test, D. W., Fowler, C. H., Richter, S. M., White, J., Mazzotti, V., Walker, A. R., and Kortering, L. (2009) Evidence-Based Practices in Secondary Transition. *Career Development for Exceptional Individuals* 32. Pg. 115-128.
- ¹³Langi, F. F. G., Oberoi, A., Balcazar, F. E., and Awsumb, J. (2017) "Vocational Rehabilitation of Transition-Age Youth with Disabilities: A Propensity-Score Matched Study." *Journal of Occupational Rehabilitation* 27.1. Pg. 15-23.
- ¹⁴Mueser, K. T., Gottlieb, J. D. and Gingerich, S. (2013) "Social Skills and Problem-Solving Training." *The Wiley Handbook of Cognitive Behavioral Therapy*. John Wiley and Sons. Pg. 243-272.
- ¹⁵Park, E., Kim, J., and Kim, S. (2016) "Meta-Analysis of the Effect of Job-Related Social Skills Training for Secondary Students with Disabilities." *Journal of Vocational Rehabilitation* 44.1. Pg. 123-133.
- ¹⁶US Department of Labor. (n.d.) "Skills to Pay the Bills: Mastering Soft Skills for Workplace Success." *Skills to Pay the Bills*. US Department of Labor. <https://www.dol.gov/odep/topics/youth/softskills/softskills.pdf> (Accessed 02/01/2016).

Culture as Resistance: A Study of the Warsaw and Łódź Ghettos

BY MIRANDA BRETHOUR, DR. JAN GRABOWSKI
UNIVERSITY OF OTTAWA

This research aims to prompt a reconceptualization of resistance, foremost challenging the simplistic approach to studying resistance in the ghettos, wherein armed revolt is the only relevant method considered. Although it is undeniable that culture in its many forms did exist prominently in the ghetto, whether that was simply whistling Beethoven during work or attending a theatrical performance, it requires a deeper analysis to understand and place it within the sphere of resistance.¹ A definition of resistance will be formulated on the basis of Nazi policy in regards to the Jewish people and supported by previous scholarly definitions. This step is integral to placing culture within the narrative of resistance, and thus achieving a more nuanced perspective. Scholars of cultural resistance have emphasized the necessity of approaching culture in the ghettos from such a perspective, as since “performance took place under various conditions, we cannot generalize about these contexts nor can we assert a single reason why an inmate created art under such circumstances.”² This quote refers to the concentration camps in particular; nonetheless, it is equally applicable to the ghettos. In spite of the complexity of this field, identifying trends within cultural resistance remains a valuable contribution to the historiography.

Cultural resistance was a common and widespread mode of resistance in the ghettos that helped the Jewish populations protest the occupation on a number of levels. After studying a selection of sources from the ghettos of Warsaw and Łódź, I was able to conclude that culture fulfilled three main roles within the ghetto, all of which fall under the umbrella of resistance to the Nazi occupation. They are as follows:

1. The production of culture as a method of physical survival
2. Culture as a vehicle of expression
3. The role of culture as psychological relief

This three-tiered approach is not to oversimplify the role of culture in the ghetto. Rather, it acknowledges the complexity of culture, yet sacrifices some of the nuance for the sake of clarity.

This article will consider both art and intellectual culture. These are often studied as separate entities; however, in the ghettos art and intellectual culture were both practiced to fulfill the same three roles illustrated above. In the larger ghettos of the Generalgouvernement, the practice of music, education, theatre, and the like, took place within an interrelated clandestine or formal system. There are exceptions, such as the individual street singers or small clandestine education groups. Nonetheless, the prevalence of webs connecting art and intellectual culture in the Warsaw and Łódź ghettos render their simultaneous study warranted.

It is imperative to note, before delving into the details of culture in the ghetto, that a certain segment of the population is missing from the narrative. Culture often overrode religious tradition, as a concert poster from the Warsaw ghetto advertises a concert planned for a Saturday, the Sabbath, and a day of rest.³ Thus, cultural activities in the ghetto took place regardless of the religious calendar. Furthermore, those who actively practiced Judaism in the ghettos faced distinct issues in addition to those faced by the non-Orthodox population. While the meager rations were a problem for many, since the meat available was horse meat or other meats not prepared according to kashrut, Orthodox Jews would have the choice to disregard their traditions or find a way of illegally procuring a substitute on the black market. In result of these additional hardships and others, few Orthodox Jews survived. This is one of the shortcomings of studying culture in the ghettos: while a substantial segment of the population can be accounted for, there are inherent limits to research.

A substantial basis of historiography exists upon culture in the ghettos that provided a reference point. The majority of these works takes a microscopic approach, addressing a single aspect of culture and analyzing its place within the ghetto. This includes a tendency to romanticize culture in the ghettos.⁴ While this metaphysical relationship with culture did exist, the primary documents prove it to be not as prominent or all-consuming as often presented in the literature. Another shortcoming of the literature is to present cultural resistance as a product of the Jewish

intelligentsia alone. Culture in the formal sector—for example musical and theatrical performances held indoors—did constitute mainly of the wealthier class in result of the admission fee. However, street performances and other mediums of informal culture represented the full breadth of society. This is necessary to mention since many scholars seem to marginalize culture on the basis that it is a production of the intelligentsia alone. Furthermore, there is a considerable lack of scholarship that tackles culture as a method of resisting physical oppression. If culture as physical survival is included, such as in Susan M. Kardos’ article on clandestine schooling as resistance in the Warsaw ghetto, it is minute in scope and dealt with as an aside.

Two books in particular move away from these patterns and were both sources of inspiration and primary material for this research. Despite their differences, Gila Flam’s *Singing for Survival* and Shirli Gilbert’s *Music in the Holocaust Ghettos and Camps* challenge the over-represented viewpoint that culture acted solely as psychological resistance alone. This research supports this perspective and goes further by classifying culture as one of the common methods of resisting the occupying regime not only in the psychological sphere, but in the physical sphere as well.

A set of boundaries for what can be qualified as resistance are needed in order to justify the above statement. Scholars seem to agree upon certain characteristics of resistance, including that it must occur within the context of oppression and can be loosely defined as “opposition to the perpetrators,” as stated by Raul Hilberg.⁵ Hilberg expands upon this definition to explain that actions such as smuggling or any other attempt at survival cannot be categorized as resistance, since they did not necessarily thwart the perpetrators. This led him to conclude in *The Destruction of the European Jews* that the Jews responded to Nazi persecution with little resistance. They reacted to decrees and orders with automatic compliance, failing to oppose the perpetrator in any form.⁶ This fails to hold true in the context of cultural decrees. Although armed revolt in the ghettos and camps was limited in sphere and success, this conclusion excludes so-called passive resistance, otherwise known as everyday resistance. The goals of the Nazi

regime in regards to European Jews must be examined to arrive at a definition of resistance that properly includes this 'passive' resistance within its boundaries. Following the functionalist line of thought, during the invasion of the Soviet Union in the summer of 1941, the Nazis formulated and begun to initiate the Final Solution, which aimed for complete physical eradication of the Jewish people and their cultural-religious traditions. This was reflected in the treatment of the Jews compared with other persecuted groups, as stated by Emil Fackenheim: "There were those whose crime was a doing—political opponents, common criminals, Jehovah Witnesses, some clergymen... And there were those whose crime was a being."⁷ During the Wansee Conference held in January 1942, high-ranking Nazi officials planned the murder of approximately eleven million Jews, including those residing in countries not under German-occupation such as the United Kingdom and Portugal.

Although complete annihilation was not necessarily a policy until the late summer of 1941, the conditions established by the Germans within the ghetto from the outset were detrimental to both the physical and psychological well-being of the Jewish population. Resistance in the ghettos therefore constituted of any effort, physical or verbal, undertaken to protest the Germany policies, including ghettoization, the ration allotments, as well as cultural and religious regulations. This definition of resistance properly includes attempts by Jews to physically survive the ghetto, maintain a sense of Jewish identity and tradition, as well as generate armed resistance, all methods of challenging the perpetrators. Scholars who do include cultural resistance within their definition of resistance often deal with it as a marginal issue, yet diaries from the period demonstrate that culture as a form of resistance was not a rare occurrence.

Michael Marrus puts forth in his historiography of resistance that the definition of resistance in the context of the Holocaust must be broad due to the diversity of circumstance faced by European Jews.⁸ This diversity of circumstance is best illustrated in Susan M. Kardos' article on clandestine schooling and resistance in Warsaw:

The Jews of the Warsaw Ghetto faced multiple struggles. There was the personal daily struggle against hunger, disease, poverty, overcrowding, and the possibility of deportation, which was intensified by the struggle to maintain a sense of dignity, normalcy, and hope in the face of brutal inhumanity. Then there was the community's struggle to stay organized and cohesive and to maintain social services,

political organization, religious life, and governance within the Ghetto walls. Finally there was the broader struggle against historical and cultural eradication and, ultimately, against complete, collective annihilation.⁹

Jews residing in the ghettos in Poland faced persecution on many levels; henceforth, a study of resistance in such a context must consider how people may have resisted against any aspect of persecution.

It is first essential to understand that policy on cultural activity was not uniform within the ghettos of the Generalgouvernement, but was instead left up to the individual German authorities to decide. In Warsaw, education was illegal until October 1941, making clandestine and semi-clandestine institutions widespread in the ghetto.¹⁰ By contrast, in Łódź, the chair of the Judenrat Chaim Rumkowski created an extensive network of schooling from primary years to university which was recognized as legal by the occupiers.¹¹ Although these two ghettos are being used as case studies in this research, they were very different in terms of cultural structure. On April 30, 1942, Heinz Auerswald, an SS lawyer, forbade the performance of Aryan literature, art, or music in the Warsaw ghetto's cafés and theatres. The order existed for almost a year prior to April 1942, according to Chaim Kaplan, but was only enforced upon Auerswald's decree.¹² In Łódź, the orchestra played works by Beethoven and other Aryan composers, suggesting either that no such ban existed, or that it was not enforced. Since the Germans allowed more freedoms in Łódź, more of the cultural institutions were run by the Judenrat and under direct supervision by the authorities compared to Warsaw.

Many cultural institutions and outlets existed in both Warsaw and Łódź. In Warsaw, YIKOR was an illegal organization created to promote the Yiddish language in the ghetto. A similar organization, Tkuma, was created to spread the Hebrew language.¹³ These groups, independent from the Judenrat and German authorities, organized many events, such as lectures and meetings, mainly directed at young people. Adam Czerniakow, the chair of the Judenrat, also took extensive efforts to create cultural life within the ghetto. He attempted to legalize the ghetto symphony orchestra when it became illegal in 1942 and also organized various performances.¹⁴

The center of Łódź's cultural web was Marysin, an agricultural area where youth took part in organized schooling and productions at the House of Culture. The system in Łódź was quite different than that of Warsaw, as cultural productions in Łódź most often took place under the watchful eye of Rumkowski, the chair of the

Judenrat. Nonetheless, there were similarities between cultural life in the two ghettos. In the performance sector, both ghettos had a legal symphony orchestra as well as theatre troops. A broad informal cultural network also existed.¹⁵ In the evenings and times after curfew, groups gathered at individual homes to exchange ideas, teach, read poetry, and more.¹⁶ Whether formal or informal, culture had practical value for a number of ghetto inhabitants. Those who possessed cultural talents used their skills as a source of income or to amass resources. Their success varied and was dependent not only on the skill of the artist, but also on the venue and time period. Due to the nature of occupation as an assault on all aspects of Jewish life—physical, cultural and psychological—these actions can and should be considered a form of resistance.

Culture as a mode of physical survival was most commonly seen in education, particularly in music and visual arts. In Łódź, many young students wrote in their diaries that one, if not the only, reason they attended school was to receive their daily allocation of soup. Dawid Sierakowiak, a young diarist from Łódź, did not dwell upon the daily meal as his central reason for attending school, but mentioned its benefits nonetheless: "At school our studying proceeds at a rapid pace. The soup they cook for us here, though not too rich or thick because of the general lack of potatoes in the ghetto, provides an excellent shot of energy during classes. After all, the long trip to Marysin wears us down terribly."¹⁷ Sara Zyskind, another student in the Łódź ghetto, made use of the ghetto schooling system as a source of food alone. In contrast to Dawid, she does not mention her classes as a source of relief from psychological stress of the ghetto. Sara depended on the rations received from school, as she and her cousin Salek, "looked forward eagerly to the beginning of the new school year, when [they] would again have enough to eat."¹⁸ When her father fell ill, Sara snuck her allotted soup and meat patty out of the school, in order to provide for her father after her mother's death.¹⁹ In an interview, Sonya, one of the teachers from the Łódź ghetto, confirmed that one of the school's major objectives was to give the students safety from the streets of the ghetto and provide them with a single meal a day.²⁰

In the fall of 1941, primary schooling for children up to the age of eleven was legalized in the Warsaw ghetto. Thousands of children made use of the school's breakfast system.²¹ Limited primary evidence exists on the use of schooling as a source of food in Warsaw, both because the legalized system was short-lived, and because the children who took part in it were quite young. In Warsaw, clandestine

schooling was more prominent, and no sources suggest that students were offered anything concrete for participating. This can explain the divergence between Warsaw and Łódź in terms of students who studied for practical gain versus psychological relief.

Like their students, educators were also motivated by the income and food they could potentially receive from their occupation. Dawid Sierakowiak was both a student and an educator in the Łódź ghetto, and he tutored many students in the ghetto for pay. He justified his tutoring work, in addition to his regular school work and manual labour, noting “cash is cash, and the main thing is to have something to eat and to survive.”²² He bartered with families for better prices, sometimes refusing to take on a student because he didn’t believe it was worth the amount their families were offering to pay: “All they would give me is 1.50 RM for six lessons a week. I want a minimum of 40 pf an hour, and I’m not going to give up even a single pfennig. I still value the little remaining energy that I have.”²³ Music teachers also relied upon their craft as a source of income in the ghetto. Władysław Szpilman and ‘Professor Kellerman,’ as he is called in Mary Berg’s diary, are two examples of highly qualified musicians who took to offering lessons in order to survive.

Artists and intellectuals who could not attend organized schools such as those in Łódź took their craft to the ghetto streets, hoping to supplement their food rations or to purchase necessities. The situation for artists and intellectuals was particularly difficult following ghettoization since, especially in Warsaw, where no established cultural venues were opened at the outset, they had lost their livelihoods.²⁴ This forced them to pursue their craft in any way that would enable them to make a living. Professor Kellerman, a violinist previously of the Leipzig Conservatory, played his violin outdoors and received pieces of bread and coins from his listeners. His talents as a musician and as a well-paid music teacher sustained him and his wife during their residence in the ghetto.²⁵

Musicians also played in cafés and restaurants for wages, to an audience which consisted mainly of the ghetto intelligentsia. The Warsaw pianist Władysław Szpilman was one such musician. During his prewar career, Szpilman performed in concert halls and on the Polish Radio. In the ghetto, he played at venues such as Café Nowoczesna on Nowolipki street to support himself and his family, adapting his livelihood to suit to the ghetto’s circumstances. He remarked in his diary, “Life, although so unimportant, had none the less forced me to overcome my apathy and seek some way of earning a living.”²⁶ Adult choirs were another aspect

of the formal culture sector in the ghetto. Although no primary documentation could be found revealing their motives, choir performers in Warsaw received benefits such as free meals and the conductor a salary of 300 zloty per month.²⁷

In his *Notes From the Warsaw Ghetto*, Emmanuel Ringelblum captures the commonality of singing in the streets for money. From the spring of 1941 to 1942, he described musicians singing in the streets alongside their children, who collected the coins.²⁸ In Łódź, Yankele Hershkowitz was the most popular street performer, and as a result, one of the few who managed to achieve enough profit through this profession to survive. He was known after every song to cry out “a new song for ten Pfennig and no more” thus, earning an income.²⁹

In various forms, culture was used in the ghetto as a method of physical resistance. Through culture, the Jews were able to attain food, medical treatment and other essentials, challenging the German policy of oppression and annihilation. Culture was one of the only ways Jews could express themselves emotionally, and personal expression offered Jews an opportunity to assert their humanity. Theatre, poetry, and singing were more than entertainment: they were modes of social commentary. They expressed resentment towards the German and Jewish authorities, thoughts of revenge, the oppression of the ghetto, and connections to Jewish tradition. Even in Łódź, where cultural activities were heavily supervised, performers managed to covertly express themselves through art. This primarily highlights the lack of passivity among the Jewish population. Jews in the ghetto were actively aware of German injustices and spoke out against them, despite personal risk, as insulting the Germans or maintaining a connection to Jewish religious identity countered German rule and was a punishable offence in the ghetto.

Janina Bauman observed a cabaret performance of “sketches of ghetto life” which exposed the corruption and indifference that characterized the ghetto.³⁰ Similarly, Mary Berg noted many instances of such criticism during theatrical performances in Warsaw. During a revue at the Femina Theatre, “there were biting satirical remarks directed against the ghetto government and its ministers,” including the chair of the Judenrat Adam Czerniakow.³¹ Unlike the symphony, there are no accounts that suggest theatre performances were suppressed for their open condemnations of ghetto authority, making them a source of popular entertainment in the ghetto. According to actors who performed in ghetto theatrical productions, the ability to be on stage and express themselves was the most important

aspect of the art.³²

Painting offered an opportunity for the artist to visually represent the ghetto how they chose. As the artist is often considered inseparable from their work, it is inevitable that visual art from the ghetto would reflect the author’s emotion. Oskar Rosenfeld contemplated the difficulty of representing the ghetto of Łódź through visual art, as an entity “outside of the realm of civilization.”³³ He noted several ghetto motifs that appear in visual art as being accurate expressions of the ghetto environment, such as people pulling carts, the constant hunger, or the corruption of the Order Service.³⁴ These ghetto motifs appear in Mary Berg’s descriptions of her art course in the Warsaw ghetto. When representing the ghetto through visual art, the students chose to present its “misery figures.”³⁵

Poetry was another method of expression. Władysław Szlengel composed the poem “Telephone” in the Warsaw ghetto, that laments the break in communication with the world outside of the ghetto:

Within my heart broken, and sick,
with my thoughts on the other side
I was sitting one evening
next to the telephone-

And I think: let me ring
someone on the other side
when I am on telephone duty
in the evening-

And suddenly I realize: my God-
there is actually no one to call,
in nineteen thirty-nine
I went on a different road.

Our ways have parted,
friendship sunk to the bottom
and now well...there is no one
I can telephone.⁵⁶

Through poetry, Szlengel expressed his animosity of the German policy of ghettoization and its personal impacts on his own life. The poem emphasizes Szlengel’s intense, yet ultimately fruitless, desire to speak with those on the other side of the ghetto wall. This resonates with the broader psychological impact of ghettoization, which uprooted and isolated Jews from their previous lives.

Singing provided the same relief as theatre, painting, and poetry. Miriam Harel, a survivor of the Łódź ghetto, stated, “Song was the only truth. The Nazis could take everything away from us, but they could not take singing from us. This remained our only human expression.”³⁷ This quote overtly points towards the role of culture—in this case singing—as a method of directly resisting Nazi policy. Culture was a way

not only of physically and psychologically resisting the ghetto, but also of directly challenging Nazi policy that treated the ghetto inhabitants as subhuman.

In Gila Flam's book on singing in the Łódź ghetto, she records many songs that showcase the role of singing as expression. "The Notorious Ghetto," composed by the infamous ghetto singer Yankele Hershkowitz, was one of such songs. Its use of political and social satire pokes fun at the systematic nature of Nazi policy and the Judenrat:

The notorious ghetto,
It runs like clockwork,
Everything is in order,
No unemployment,
They eat roast,
Play cards,
The entire Beirat is corrupt.³⁸

Hershkowitz was well known for his risqué songs that critiqued the agents of control in the ghetto. The most remembered by survivors was his song about Chaim Rumkowski. When a policeman attempted to arrest Hershkowitz for insulting Rumkowski, proof that ridiculing ghetto authority carried consequences, the crowd surrounded him and enabled him to escape arrest.³⁹ Oskar Rosenfeld also mentioned Hershkowitz's street performances in his diary, recalling a particular song that made fun of the Germans and their girlfriends.⁴⁰ He sung this tune directly within range of the Germans themselves, refusing to relinquish his human ability to express himself in the face of threat.

A popular song of the Warsaw ghetto, "Money, Money," set to the tune of a pre-war American jazz song, closely reflects the content of "The Notorious Ghetto" and other songs from Łódź:

Money, money, money is the best thing,
The Jewish policeman is just a scoundrel,
Puts you on the train and sends you
away to a camp.
Money, money, money is the best thing.⁴¹

As this excerpt suggests, the song highlights the corruption and disparity between the poor and elite in the ghetto. It particularly criticizes the Jewish Order Police for its collaboration with the Nazis. The Jewish Order Police was forcibly created for the purposes of facilitating the compliance of the Jewish population, enforcing Nazi decrees and assisting with deportations to the labor and death camps.

The call for revenge was also a common topic expressed in songs. This could take the form of a political prophecy, that the Jews would live through this period to enact revenge upon their oppressors. Sometime after the Aktion of July 1942, Władysław Szpilman remembered walking with his work group through the Warsaw ghetto on New Year's Eve. When asked to sing by the

Germans, they sang the Polish patriotic song "Hey, marksmen arise!" in a place when declarations of Polish patriotism were banned.⁴²

Not only was culture a medium through which to express discontent, frustration, and revenge, but also a way to evoke connections to one's Jewish identity. This aspect of expression somewhat merges with the third element of culture in the ghettos; the idea of culture as morale sustenance. However, the link between culture and identity will be understood as expression for the purpose of this research due to the reliance on verbal expression—lyrics—to form this connection.

In the Warsaw ghetto, the Germans forbid religious study, teachings, and public or private worship to take place.⁴³ With so many restrictions, maintaining a connection to Jewish culture would reinforce a sense of communal belonging and identity. Pinchas Saar, the official theatre painter and set decorator in the Łódź ghetto, is an antithesis in the context of this article. He remarked in an interview with Gila Flam that he did not believe that culture, specifically theatre, was as important in the ghetto as it is often presented today; nonetheless, he ultimately recognized that the bond theatre performances were able to form with Jewish culture.⁴⁴ Since Nazi policy directly attacked Jewish traditions themselves, maintaining connection to Jewish identity through culture qualifies as resistance.

This connection was established through songs that focus on Palestine, and the future of a Jewish community there. One was composed by Yankele Hershkowitz, who described the communal desire to live in Palestine, and celebrations that would take place after the state of Israel was declared: "I'm going to Palestine, That is a golden land... Rumkowski and his army want to enter Eretz-Yisrael too. We'll make noise, All through that time, we'll have a celebration, all right!"⁴⁵ In the Warsaw ghetto, a similar song was sung by Janina Bauman's sister Sophie. Composed in Hebrew rather than Yiddish, the main idea of this song mirrors that of the Łódź ghetto. It expresses the desire of the ghetto's young population to leave the humiliation of the ghetto to build the new homeland in Palestine.⁴⁶ Therefore, culture was used as one of the "weapons of the powerless" in the ghetto, a method of expression that challenged the German notion of Jewish inferiority and provided relief from the tension of ghetto life.⁴⁷

Culture also provided a sense of normalcy or figurative escape from the ghetto, thus acting as psychological resistance to the occupiers. A select number of survivors from Warsaw and Łódź credited this to be a contributing factor to their survival, for various reasons including

a reminder of how life existed outside the ghetto boundaries. Especially for children, schooling in the ghetto played this role, inputting a sense of community and future goals in their lives. In both the ghettos and camps, scholars and survivors alike have spoken upon the importance of this drive to live as a determining factor of survival itself.⁴⁸ Those who relinquished their desire to live were known as a *muselmann* in the camps. Although they had not yet died, the inmates knew that once that stage of apathy was reached, their death was inevitable. For some, culture became this source of moral stamina that withstood against the Nazi policies.

Dawid Sierakowiak was one of the many children and young adults in Łódź who attended the school system run by the Judenrat. From his diary, it appears as though his education injected a degree of normalcy and objective in his life. Unlike other attendees, such as Sara Zyskind, he did not focus upon the daily rations as the single reason to walk to Marysin for schooling. He often mentioned his desire to learn, particularly languages, for future benefits. When he was unable to go to school due to illness, he lamented: "Damn the times when I complained about getting up in the morning and about tests. If only I could have them back!"⁴⁹

Educators at the time held different explanations for why Jewish youth continued to study in the ghetto. Two dichotomous groups exist: while one particular teacher believed they studied to satisfy an inner desire without thought of future benefit, another asserted that they attended school out of obligation.⁵⁰ Whether one group makes up the majority of ghetto students is not of importance, since primary evidence from diaries confirms that for certain youth, including Dawid Sierakowiak, studying was undertaken out of personal drive as "the miraculous way of getting away from reality."⁵¹

Survivors Vladka Meed and Janina Bauman both recalled positive memories of their ghetto educations in Warsaw. In the winter of 1941, a group of young adults, including Meed, met to listen to a speech about the Jewish writer I.L. Peretz. Afterwards, they spread to various houses to repeat the lecture. She did not recollect the discussion among the forty inhabitants of the house behind the blacked-out windows, but rather the "wonderful atmosphere, the feeling of being able, even for a short time, to get away from the bitter ghetto reality."⁵² Bauman partook in a similar system, continuing her studies with nearby friends. They established small groups led by teachers from a prewar grammar school, and walked to each other's homes every day to study subjects such as classics and mathematics.⁵³

Mary Berg was an attendee of many ghetto education initiatives. A recurring topic in her diary is a graphic art course, wherein the students learnt a range of topics from art history, architecture, and drawing techniques. Berg partook in competitions and symposiums with other students, and describes in detail her fellow classmates and the pieces they created. Her tone suggests that she derived substantial enjoyment from this course, particularly in an entry on February 25, 1941: "The atmosphere is pleasant. I feel as though every day I am visiting another world for a few hours, a world far removed from the ghostly life of the ghetto."⁵⁴

Some ghetto inhabitants clung to culture as a remnant of their prewar lives. Władysław Szpilman's family did so during the Aktion period to maintain a sense of normalcy: "My parents, sisters and brother knew there was nothing they could do. They concentrated entirely on staying in control of themselves and maintaining the fiction of ordinary daily life. Father played his violin all day, Henryk studied, Regina and Halina read and Mother mended our clothes."⁵⁵

While up to this point culture has been shown as a medium of psychological resistance, both through its potential to connect the ghetto inhabitants with their prewar-war time habits as well as providing a sense of metaphysical escape, this section will focus entirely on the latter. In the ghetto, performance became the most common method to achieve metaphysical escape. Both ghettos had theatre troops, symphony orchestras, choirs and other forms of performances. However, it is essential to acknowledge that such events were largely restricted to the upper classes of society, due to an entry fee. The elitist character of these performances appears in writings and revues from the time, including Dawid Sierakowiak's diary. He depicted one of the weekly orchestras concerts at the Cultural House in Marysin: "Today I went to the concert on Krawiecka Street again. It was the first concert worth seeing in the ghetto: a Beethoven evening. The whole of select Society gathered, bloated and dressed up."⁵⁶ His use of the word "select" to qualify society, suggests the limited social representation in the audience of formal performance. Nonetheless, this does not render the moral gains from such events meaningless.

One of the main established theatres in the Warsaw ghetto was the Eldorado on Dzielna Street. Various revues describe the psychological relief afforded by these performances, seemingly a joint product of the physical and metaphysical atmospheres. *Gazeta Żydowska* reported that a few hours of being in a warm hall in a "truly Jewish atmosphere" of Yiddish speech and music

allowed this experience to bring relief.⁵⁷ A revue from January 1941 of a performance described the transition from the beginning to end of a performance feeling alike to the end of a storm when "the sun comes out. The mood lightens. The stage is full of life, the whole ensemble sings and dances."⁵⁸

Oskar Rosenfeld, one of the contributors towards writing a chronicle of the Łódź ghetto at the time, made similar remarks about the House of Culture in Łódź. Although it was restricted in some ways due to the heavy supervision by the Nazis and Chaim Rumkowski—it has been said that the chair of the Judenrat attended almost every performance during the House of Culture's existence—the audience still managed to attain a sense of relief from the performances. In Rosenfeld's opinion, the institution itself met the metaphysical needs of Jews, and was proof the ghetto could not break Jewish resistance.⁵⁹ He believed that this institution of culture was successful at providing "400 downtrodden souls with two liberating hours" thus, bringing "honor to its name and mission."⁶⁰

Outside of these official institutions, culture was still nevertheless a source of psychological relief. Mary Berg experienced this sense of psychological escape from listening to Professor Kellner play violin outside her home at 41 Sienna street: "When he begins to play, windows open on all the floors. I often close my eyes and imagine that I am attending the concert of some great virtuoso, discreetly accompanied by a distant orchestra."⁶¹ Listening to a work of music or viewing a piece of art carried the ability to distract certain ghetto inhabitants from the reality of their situation.

One survivor connected culture, and the psychological relief attained from it, directly to her survival. When Rumkowski ordered the youth group Hakhsharah closed at the end of 1941, group members continued meeting in private. One these members was Leah, whose testimony centers around the singing that occurred at these meetings and the psychological relief from the ghetto they provided:

We did not give up singing. It was singing for its own sake. We sang all kinds of songs. Actually, we did not have any good news to talk about. We tried to forget the bad times, so we sang. It worked wonderfully! I think it was one of the things which helped us to survive.⁶²

Leah's experience does not represent the majority, however. Although this response to culture did not necessarily promise a greater chance at physical survival, it is nevertheless significant as a psychological method of resisting the German policy of the ghetto.

This reflects Czerniakow's response to criticism of his focus upon creating cultural activities for the ghetto population, and the youth in particular. His intent in organizing these activities was, in his view, comparable to a captain on a sinking ship that requests jazz music be played while the ship is going under: simply to raise spirits.⁶³

In the ghettos, cultural activity played a role as psychological resistance. Ghetto inhabitants practiced culture as a way of remaining connected with their prewar habits, and metaphysically escaping the oppression created by German policy. In his article, "Resistance as Happiness," David R. Blumenthal takes a similar approach, arguing that any attempt to fight Nazi oppression by remaining optimistic through culture, religion, and other mediums, is resistance.⁶⁴

Many barriers were overcome and risks taken in pursuing Jewish cultural life in the ghettos. It is significant to mention these since it is assumed, especially in places like Łódź, where the majority of cultural life was legal, that little difficulty stood in the way of maintaining a sense of culture. Understanding cultural activity as something that people strove to pursue despite known risk and obstacles reinforces its place within the sphere of resistance.

In Łódź the symphonic orchestra, according to a select number of reviews, performed at a high level despite numerous difficulties. Oskar Rosenfeld noted that the conductor Theodor Ryder managed to conduct the orchestra with no full score, and not all the instrumentation sketched.⁶⁵ The struggles to perform their weekly Wednesday evening concert expanded over time, as various members (including the concertmaster) had been deported. Despite this, concerts continued up until January 17, 1944, when the Germans confiscated musical instruments and abruptly brought an end to musical performance in the Łódź ghetto.⁶⁶ Rosenfeld also remarked in his diary that playing in this orchestra was not the job of the members. They would most likely work, doing physical labor during the day, and would then play in the orchestra without any pay at night.⁶⁷

The symphony in Warsaw, which became legal in the beginning of 1941, underwent similar struggles. When the orchestra was suspended for playing German music on April 11, 1942, musical life continued underground.⁶⁸ Small groups were organized to perform and listen to music in peoples' homes. Furthermore, although the Germans had forbidden non-Jewish music, this continued even after the orchestra was forcibly disbanded for violating that decree. The Polish composer Chopin was one of the forbidden artists; however, musicians

continued to play his lesser-known pieces, and, if questioned, lied and said it was another composer.⁶⁹ Legality did not influence music in the informal sector. Janina Bauman remembered listening to the one record she had, Beethoven symphony no. 5, on the gramophone during the cold night after curfew with her friends.⁷⁰ Most clandestine or informal cultural activities took place after curfew, which at the beginning of the occupation in Warsaw was from seven at night to five in the morning. During this time people gave recitations, and actors and musicians were known to perform.

Theatre performances in Warsaw faced numerous difficulties, including a lack of materials. An amateur theatre group in early 1940 made a curtain out of a sheet, with decorations made out of tablecloths and bedspreads. Performances were held under carbide lamps when the electricity went out, and the audience wrapped themselves in blankets while the performers on stage went blue with cold.⁷¹

Especially in the Warsaw ghetto, where no extensive legal education network existed, education was a risky endeavor. Two examples from the ghetto in Warsaw illustrate the drive of young people to find a way around the barriers and continue their intellectual lives, as well as the perseverance of the parents to create an education for their children. Frieda Aaron, who completed years five and six as well as two years of gymnasium underground, spoke of the dangers associated with pursuing clandestine education in the Warsaw ghetto during her interview with the USC Shoah Foundation. She and her friends had to carry their books through the ghetto streets under their coats, in full knowledge that if they were caught by the authorities, they would have been killed immediately.⁷² Vladka Meed observed her neighbour hurrying her daughter to similar secret classes.⁷³

Education did not carry the same risk in Łódź, but difficult circumstances existed nonetheless. Dawid Sierakowiak often complained of how difficult it was to focus on his studies and learn new ideas due to the hunger. He also had to walk to Marysin, far from where he lived in the ghetto, in order to attend classes. With the extreme hunger and illness prevalent in the ghetto, these two seemingly simple tasks would have required incredible effort. In light of the perseverance to continue cultural life despite difficulties and the ways in which it was used, the production of culture in the ghetto can be considered an act of resistance.

Culture was also practiced as a form of collaboration with the German occupiers. This lies outside the sphere of resistance, but is imperative to note nonetheless. Culture

is also essential to challenge the idea that the Jews were utterly complacent to the German occupiers. Though small and ultimately fruitless at bringing about change, these everyday acts of cultural resistance combated the ideas of the Nazi regime by fighting to survive and actively rejecting German policies in all spheres. Thus, the definition of resistance in the context of Holocaust must be expanded to include these acts of cultural resistance, lest the full breadth of resistance historiography remain incomplete.

as a form of collaboration in the ghettos is difficult to identify; one must attempt to unearth and interpret the intent of past individuals from primary documents as closely to reality as possible while avoiding conclusions based on moral bias. Where this phenomenon is most easily recognized is within the ranks of the Jewish authority figures, the Judenrat, and Jewish Order Police. The creation of cultural activities for the ghetto population by the Judenrat, for example Chaim Rumkowski's activities and schooling for children, can be viewed as ultimately assisting the Germans in the deportation and murder. Culture potentially calmed the population, reminding them of times before the ghetto and thus making them less fearful and suspicious of what was to come. This calming and nostalgic quality of culture dichotomously could be used both by the population as a method of psychologically resisting German policy, whilst also being a way the authorities could render the populations more complacent to orders.

Czerniakow's quote, that his intent in organizing these activities was comparable to a captain on a sinking ship who requests that music be played, can also be understood as a form of collaboration. Through culture, he attempted to shield the population from the reality of their fate. This reassurance would hinder the development of physical resistance, whether that be by arming themselves, or attempting escape. He created an orchestra within the Order Service that played at celebratory events, such as the opening of a new playground for children. The chairs of the Warsaw and Łódź Judenrat, Rumkowski, and Czerniakow used culture in a way that can be interpreted as supporting the Germans.

Placing culture in the ghetto against the backdrop of Nazi policy is integral to understanding its nature. The Nazis attacked all spheres of Jewish life—physical, cultural, and religious. Even before the establishment of the Final Solution, the Jews were physically abused and humiliated by the perpetrators. In this context, Jews in the ghettos turned to productions of culture that promised a chance at both physical and psychological survival. Whether it was used as a means of making money, expressing suppressed feelings about the ghetto and related policies, or psychological relief from the oppressing ghetto life, culture became an everyday mode of resistance against the occupying forces and the policies they had imparted.

A significant amount of research could still be done on this subject, as the cultural life of each ghetto and camp are distinct, shaped by countless factors, including the nature of the prewar cultural life and occupying authorities in the area. However, this topic

is also essential to challenge the idea that the Jews were utterly complacent to the German occupiers. Though small and ultimately fruitless at bringing about change, these everyday acts of cultural resistance combated the ideas of the Nazi regime by fighting to survive and actively rejecting German policies in all spheres. Thus, the definition of resistance in the context of Holocaust must be expanded to include these acts of cultural resistance, lest the full breadth of resistance historiography remain incomplete.

References

- ¹Kovaly, H. (1986) *Under a Cruel Star: A Life in Prague 1941-1968*. New York: Holmes & Meier. Pg 4.
- ²Rovit, R. (2005) "Cultural Ghettoization and Theatre during the Holocaust: Performance as a Link to Community." *Holocaust and Genocide Studies*. 19. 3. Pg 476.
- ³Engelking, B. (2008) *The Warsaw Ghetto: A Guide to the Perished City*. New Haven: Yale University Press. Pg 579.
- ⁴Adler, E. (2006) "No Raisins, No Almonds: Singing as Spiritual Resistance to the Holocaust." *Interdisciplinary Journal of Jewish Studies*. 24. 4. Pg 50-66.
- ⁵Gottlieb, R. (1983) "The Concept of Resistance: Jewish Resistance During the Holocaust." *Social Theory and Practise*. Pg 32.
- ⁶Hilberg, R. (1985) *The Destruction of European Jewry*. London: Holmes & Meier. Pg 293.
- ⁷Fackenheim, E. (1982) "The Spectrum of Resistance During the Holocaust: An Essay in Description and Definition." *Modern Judaism*. 2. Pg 119.
- ⁸Marrus, M. (1995) "Jewish Resistance to the Holocaust." *Journal of Contemporary History*. 30. Pg 83-110.
- ⁹Kardos, S. (2002) "'Not Bread Alone': Clandestine Schooling and Resistance in the Warsaw Ghetto during the Holocaust." *Harvard Educational Review*. 72. 1. Pg 36.
- ¹⁰Engelking, B. Pg 538.
- ¹¹Horwitz, G. (2008) *Ghettostadt: Lodz and the Making of a Nazi City*. Cambridge: Belknap Press. Pg 80.
- ¹²Kaplan, C. (1973) *Scroll of Agony: The Warsaw Diary of Chaim A. Kaplan*. Bloomington: Indiana University Press. Pg 322.
- ¹³Engelking, B. Pg 6, 538.
- ¹⁴Czerniakow, A. (1999) *The Warsaw Diary of Adam Czerniakow*. Chicago: Ivan R. Dee. Pg 358.
- ¹⁵Zelkowitz, J. (2002) *In Those Terrible Days: Notes from the Lodz Ghetto*. Jerusalem: Yad Vashem.
- ¹⁶Aaron, F. (2015) Interview 802, Visual History Archive. USC Shoah Foundation.
- ¹⁷Sierakowiak, D. (1996) *The Diary of Dawid Sierakowiak: Five Notebooks from the Lodz Ghetto*. Adelson, A. New York: Oxford University Press. Pg 87.
- ¹⁸Eilenberg-Eibeshitz, A. (2000) *Preserved Evidence Ghetto Lodz, Volume II*. Israel: H. Eibeshitz Institute for Holocaust Studies Haifa. Pg 561.
- ¹⁹Ibid. Pg 556.
- ²⁰Bendremer, J. (1991) "Female Survivors of the Holocaust: Heroines All." *Educational Gerontology*. 17. 2. Pg 87.
- ²¹Hilberg, R. (1992) *Perpetrators, Bystanders, Victims: The Jewish Catastrophe 1933-1945*. New York: Harper Collins. Pg 144.
- ²²Sierakowiak, D. Pg 115.
- ²³Ibid. Pg 88.
- ²⁴Engelking, B. Pg 531.
- ²⁵Berg, M. (2006) *The Diary of Mary Berg*. Oxford: Oneworld Publications Limited. Pg 63.
- ²⁶Szpilman, W. (2002) *The Pianist: The Extraordinary Story of One Man's Survival in Warsaw, 1939-1945*. London: Orion Books Ltd. Pg 11.
- ²⁷Kermish, J. (1999) *To Live With Honor, To Die*

- With Honor. Jerusalem: Yad Vashem. Pg 355.
- ²⁸Ringelblum, E. (2006) Notes from the Warsaw Ghetto. New York: iBooks inc. Pg 283.
- ²⁹Flam, G. (1992) Singing for Survival: Songs of the Lodz Ghetto, 1940-1945. Chicago: University of Chicago Press. Pg 25.
- ³⁰Ibid. Pg 63.
- ³¹Berg, M. Pg 102.
- ³²Rovit, R. Pg 464.
- ³³Rosenfeld, O. (2002) In the Beginning Was the Ghetto: Notebooks from Lodz. Evanston: Northwestern University Press. Pg 238.
- ³⁴Ibid. Pg 238, 240.
- ³⁵Berg, M. Pg 70.
- ³⁶Szlengel, W. "Telephone," Trans. Carpenter, J. and B.. Chicago Review.
- ³⁷Flam, G. Pg 1.
- ³⁸Ibid. Pg 73.
- ³⁹Ibid. Pg 25.
- ⁴⁰Rosenfeld, O. Pg 273.
- ⁴¹Gilbert, S. (2005) Music in the Holocaust: Confronting Life in the Nazis Ghettos and Concentration Camps. Oxford University Press. Pg 23.
- ⁴²Szpilman, W. Pg 125.
- ⁴³Dawidowicz, L. (1976) The War Against the Jews 1933-1945. New York: Bantam Books. Pg 335.
- ⁴⁴Flam, G. Pg 131.
- ⁴⁵Ibid. Pg 77.
- ⁴⁶Bauman, J. (1986) Winter in the Morning: A Young Girl's Life in the Warsaw Ghetto and Beyond 1939-1945. London: Virago Press. Pg 86.
- ⁴⁷Dawidowicz, L. Pg 294.
- ⁴⁸Levi, P. (1958) Survival in Auschwitz. New York: Touchstone.
- ⁴⁹Sierakowiak, D. Pg 37.
- ⁵⁰Kermish, J. Pg 502.
- ⁵¹Ibid. Pg 503.
- ⁵²Bromberg, B. and H. (2011) Portraits in Literature: The Jews of Poland. Portland: Vallentine Mitchell. Pg 218.
- ⁵³Bauman, J. Pg 39.
- ⁵⁴Berg, M. Pg 42.
- ⁵⁵Szpilman, W. Pg 94.
- ⁵⁶Sierakowiak, D. Pg 174.
- ⁵⁷Engelking, B. Pg 557.
- ⁵⁸Ibid. Pg. 556, 557.
- ⁵⁹Rosenfeld, O. Pg 68.
- ⁶⁰Ibid. Pg 81.
- ⁶¹Berg, M. Pg 62.
- ⁶²Flam, G. Pg 156.
- ⁶³Czerniakow, A. Pg 376.
- ⁶⁴Blumenthal, D. (2004) "Resistance as Happiness." CrossCurrents 64. 1
- ⁶⁵Trunk, I. (2006) Lodz Ghetto: A History. Bloomington: Indiana University Press. Pg 380.
- ⁶⁶(1984) The Chronicle of the Lodz Ghetto 1941-1944. Dobroszycki, L. Binghampton: Vail-Ballou Press. Pg 471.
- ⁶⁷Rosenfeld, O. Pg 80.
- ⁶⁸Czerniakow, A. Pg 342.
- ⁶⁹Ibid. Pg 579.
- ⁷⁰Bauman, J. Pg 54.
- ⁷¹Engelking, B. Pg 532, 555.
- ⁷²Aaron, F. Interview 802.
- ⁷³Meed, V. (2015) Interview, Visual History Archive, USC Shoah Foundation.

Troubling the Nation: Crafting the New Gay in a Post 9/11 Environment

BY TARA A. CASEY, MARTHA J. GARNER, REBEKAH S. KAMP,
& GABRIEL VILLARREAL
NORTHERN ARIZONA UNIVERSITY

In the years following September 11, 2001, the United States witnessed drastic shifts in the politics of belonging to the nation. This generated a heightened sense of nationalism that was inclined to reject brown-skinned peoples, and simultaneously coincided with an increased push for “gay rights.” This created an intersection of sexuality and citizenship that operated to redefine notions of queerness within the United States. As a result, what emerged at this point in time was a unique national and cultural project that we will refer to throughout this paper as “crafting the new gay.” Through the use of the phrase “crafting the new gay” as a nationalist project of modernity, we wish to reference the similarly structured project set forth in 1998 in Afseneh Najmabadi’s “Crafting an Educated Housewife in Iran.”¹ This project by the state of “crafting the new gay” sought to incorporate certain queer bodies into the nation as a marker of modernity. In our exploration of this narrative created by the nation, we set forth to answer the following questions: since belonging to the nation was redefined, how were certain bodies brought in at the same time that other bodies were pushed out? Furthermore, how did this reorganization within the state of bodies perpetuate U.S. exceptionalism, both from a national and transnational standpoint? In this paper, we will argue that the U.S. in the post 9/11 era proliferated fear and anxiety in regards to the presence of brown and specifically Muslim bodies, and in turn replaced those bodies with homonormative subjects who were deemed to be more valuable to the nation at this point in time. This was accomplished in part through the privileging and prioritizing of the “rights” of white, gay bodies while simultaneously neglecting the hostile environment directed at brown bodies. White U.S. homosexuals were selectively marketed to represent the embrace of progressive sexual politics, with the intent to foster heightened patriotic sentiments designed to recover the American nation’s mental and economic health. Modalities of consumerism, media representations, and legislation were deployed to help generate the acceptance of homonormative subjects under the guise of modernity.

We will build our case through an exploration of what belonging to the nation means by critiquing the selective manner in which the status of citizenship was simultaneously bestowed on some and denied to others, perpetuating the discourse of “us versus them” by privileging bodies and feelings that align with the nation’s ideals and excluding other bodies and feelings labeled as sources of fear or threats to the state. Through this selective inclusion, the U.S. created a narrative of progression, modernity, and “gay-friendliness” that is used to “otherize” nations which Western ideals portray as backward and homophobic. However, the queer subjects embraced and celebrated by the nation were exclusively white, gay males who represented normative standards of class and race. Although this phenomenon created a façade of acceptance, there continues to be repercussions for non-hegemonic members of the queer community as well as brown and Muslim bodies that have been systematically excluded.

Selective Citizenship

When examining the political workings behind welcoming queer bodies into the nation’s hegemony, there were extreme constraints put on this acceptance in order to adhere to the “ideal citizen” model developed by the state, resulting in conformity to the state’s identity. The acceptance of normative, white, cisgender bodies occurred through their own assimilation and the othering of Muslim and brown bodies. This transformed these undesirable brown bodies into the new target of the nation, enabling a platform for certain queer bodies to gain support within the nation after the events of 9/11 and the War on Terror that soon followed. In the words of feminist theorist Jasbir Puar, even though “patriotism during the post 9/11 crisis was inextricably tied to a reinvention of heterosexual norms for Americans, progressive sexuality was championed as a hallmark of US modernity.”² These seemingly progressive sexual politics were presented as the new marker of national identity.

Stemming from 9/11, we saw the creation of entities such as Homeland Security, which only further entrenched a national dialogue of “us versus them.” The central focus of

this nationalist narrative was on protecting the United States from “terrorism,” which was newly identified as being embodied by brown and Muslim subjects. In the post 9/11 era, Islamophobia came to replace homophobia as a primary marker of national otherness. This was an interesting shift that occurred within the national script of belonging, whereby white gay bodies were conferred the legitimacy of citizenship, and in turn they were expected to uphold a commitment to capitalist consumption as a patriotic duty to help uplift the nation.³ It is in the context of this shift that we begin to witness the commodification of specific gay bodies with the intent to restore the nation’s financial wellbeing. Gay subjects were now seen as the repositories of modernity and progressive politics within the United States.

In addition, the state’s recognition and validation of normative, gay citizens is accomplished through the manipulation and promotion of feelings that align with the state’s agenda of “us versus them.” As part of this agenda, the nation portrays brown and Muslim bodies as objects of fear that pose a threat to the nation and its citizens. Stemming from this xenophobia, the nation supported this blatant racism by only addressing the emotions of white bodies. This simultaneously disregarded and dismissed the fears of subaltern subjects who were facing an increasingly violent response to their existence by being labeled as outsiders of the nation. In the narrative of fear perpetuated by the state, not only were brown bodies seen as targets, but white citizens were also expected to engage in “a strategy of ‘governing through neurosis’ in which the ‘neurotic citizen’ was encouraged to have a number of anxieties, including ‘about the Other.’”⁴ In this way, the feelings of the “neurotic citizen” (anxious white bodies) were validated, and in some instances, applauded by the state.

This selective engagement of emotions between the citizen and the state created larger agency for those recognized as ideal citizens, further bolstering the imperialist and dominating identity of the United States. It is important to consider that homonationalism is not exclusive to the nation, but is a product of U.S. imperialism that has spread far beyond

Western borders. The process of bolstering the state's status of sovereignty and progress (or illusion thereof) through the presence of selective homosexual acceptance and rights created benefits for the state. This tactical inclusion of previously marginalized subjects works to distract attention from the nation's perpetuation of settler colonialism, racism, acts of war, and so on, under the guise of sexual modernity. As stated by Sara Ahmed in her discussion of gay imperialism, "it is as if there is a script that is written in advance; it is as if the very point of the script is to block the critique of racism from getting through."⁵ This inclusion of white, gay bodies is used to distract from the exclusion of brown, Muslim bodies.

Although the nation is including and excluding certain individuals from the nation, the individuals who are being excluding for the nation's gain are the same individuals who are creating action against this dynamic. Agency can be seen within theorists who are supplying these essential works that spark the discussion around exploitation of Muslim individuals. Scholars such as Jasbir Puar, Sara Ahmed, and Chandan Reddy not only have been some of the first in illuminating these contradictions that exist within the politics of belonging, but they also have been extremely critical in its functioning and the danger it presents. We can find this in not only in the work of the theorists mentioned above, but also in organizations run by people of color like *Against Equality*, who are actively critiquing queer politics in the United States.

Sexual Modernity and (Re)Imagining the Nation

Although the promotion of "gay marriage and gay-friendliness is fast becoming a marker of modernity, progress, and western liberalism,"⁶ it can be viewed as a divergent tactic to what harmful acts the nation is endorsing. After 9/11, the United States boosted its commitment to capitalism and looked for subjects who could bolster capitalist consumption. Therefore, the white, cisgender, upper/middle class, gay male was targeted and given certain citizenship "rights" that they had been previously denied. With white gay male inclusion came marketing schemes and lines of products that ironically reinforced heteronormative ideals in the West. Monogamous marriage, for instance, was reinstalled as a foundational institution framing U.S. modernity. This normalization of monogamous gay marriage was absorbed into a wider capitalistic program as a way of imposing these Western ideas onto other nations. The inclusion of certain institutions, such as gay marriage, into the national discourse of belonging perpetuated the erasure of the violence and degradation

experienced within the border of the U.S. by subaltern queer subjects.

Puar asserts in her article, "Abu Ghraib: Arguing Against Exceptionalism," that the actual events that took place within the Bush administration did not reflect U.S. modernity. Puar writes, "as George W. Bush stated of the abuse at Abu Ghraib, days after the photographs had been circulating among foreign press: 'Their treatment does not reflect the nature of the American people.'⁷ Puar argues that Bush's affirmation that American values are not "perverted" or capable of doing harm, are in fact fallacies he preached to citizens within national borders and to the wider global community. Ahluwalia makes a note that Western nations use gay politics to (re) present themselves as progressive in regards to sexual politics, which they accomplish by contrasting themselves with the "illiberated" other.⁸ By marking the brown male body as effeminate and shamed, white, U.S., queer identities could then move from the national margins onto its center. In other words, the social acceptance of specific gay bodies, the internalization of Islamophobia, and the marginalization of non-hegemonic members of the queer community together enabled the promotion and advocacy of the U.S. being a "gay-friendly" nation.

Diluting Queerness

It is important to understand what homonormativity is when describing the specific queer bodies that are welcomed into the nation. Originally, as articulated by Lisa Duggan, homonormativity is a set of "politics that does not contest dominant heteronormative assumptions and institutions but upholds and sustains them while promising the possibility of a demobilized gay constituency and a privatized, depoliticized gay culture anchored in domesticity and consumption."¹⁰ Within homonormativity, certain queer subjects are assimilated into the dominant characteristics of heteronormativity, only straying from these prescribed roles in that they are, in fact, homosexual subjects. Adherence to domesticity and capitalist consumption are foundational aspects of homonormativity; through domesticity and compliance with capitalist systems, certain queer subjects are able to conform to hegemonic standards of heteronormativity, thus normalizing and mainstreaming their queerness.

Selling the New Gay

As definitions of citizenship have been broadened to include queer subjects, it is important to note that they represent homonormative notions which are predominantly white, gay, affluent,

cisgender males who are understood to be conventionally attractive within the dominant Western standard of beauty. These bodies are represented as ideal subjects both for consumption and as national citizens. Although queerness is vastly diverse, this homonormative subject eventually became the hegemonic representation of the queer community. According to GLAAD's *Where We Are 2015* report, only 4% of characters represented on television fall into the category of LGBT. Of this, only 28% of these characters are people of color; breaking this down 28% of the 142 characters leaves us with less than 40 characters representing queer people of color on mainstream television.¹¹ This homogenous identity of the queer community is constituted largely through marketing strategies, such as print ads and commercials. The last few years have witnessed an increase in corporations using queer bodies in their ads as a form of inclusion.¹² However, bodies largely represented in this marketing have been white, gay, affluent, cisgender males, as opposed to other subaltern queer subjects. Yet despite this lack of diversity, corporations continue to use queer bodies in their advertisements to tap into the queer consumer market and to expand their base for potential customers. According to Rivendell Media (2015), about 23 million dollars were spent on major ad categories in 2015.¹³ Including queer (specifically normative queer) characters in advertisements promotes an image of modernity for the company without necessarily challenging institutional oppressions that some members of the queer community experience, and in some instances reinforcing subaltern exclusion. Two examples of such homonormative advertisements that we will discuss below are an Amazon Kindle commercial and a Miller Lite print ad.

This Amazon Kindle commercial shows a white man on the beach who is unable to read the book on his tablet because of a glare from the sun.¹⁴ The white woman next to him informs him of the Amazon Kindle, which prevents glare, and he promptly buys the Kindle. To celebrate his new purchase, the man suggests that the two get drinks. The viewer is meant to believe that the man is attempting to be flirtatious, because the scene is read as a typical exchange between two heterosexual individuals within a dominant heteronormative society. In response, the woman claims her husband is getting her a drink in order to stop his advance. At the climax of the scene, the man states, "So is mine," and the camera pans over to show two men chatting at the bar. From the start, the gay bodies in the commercial are extremely underplayed and

their queerness is not revealed until the end. The representation of the two queer characters is also homonormative as they are both attractive, married, white, cisgender males who have enough disposable capital to buy a Kindle on a whim while vacationing at the beach. As white, gay males have the most agency within the queer community, they are able to expend the most financial resources to support the nation's capitalistic agenda. Therefore, these identities are selected as representatives of queerness within media advertisements.

In the 2012 Miller Lite print ad, there are again two white, attractive males who are drinking Miller Lite beers.¹⁵ The man on the right has his arm around the other, as a large Miller Lite is seen in the background with the phrase "great to see you out" above it. The ad portrays an oversaturated image of white gay males. These men are presented as potential national citizens whose whiteness, masculinity, and class status are non-threatening to the nation. These men are also subscribing to hegemonic ideals of masculinity in the way they are dressed and consuming beer. This advertisement is not overtly expressing queerness; rather, it is letting the audience know of their supposed sexual orientation through a subtle notion (as if the advertisement is saying, "P.S. they're gay" despite their other normative markers). Another problematic aspect of this ad is that it reinforces the notion of 'coming out' as a typical or singular process in the queer experience. While some may claim that these advertisements are just an outlier, there has been much appreciation for these ads among some sections of the queer community. This Miller Lite ad was nominated by the GLAAD Media Awards in Advertising in 2010 for Outstanding Print Campaign, which demonstrates the support from the gay community as they claim this to be a progressive and "outstanding" ad.¹⁶ However, reading this as an outstanding advertisement not only negates the heterogeneity within and among queer identified communities, but it also seeks to replace it with a homogenized and homonormative non-threatening image.

Commodification & Co-optation of Queer Issues

Although media representations and advertisements contribute to the normalization of gay bodies, legislation and co-opting of queer issues also aid in the project of crafting the new gay under the façade of modernity. Specifically, the commodification and co-optation of the gay marriage movement restricts understandings of queer issues and severely limits the prominence of activism outside of these issues, portraying a singular homogeneous image of queer identity

politics. In the post 9/11 era, a capitalist push for national spending within borders was framed as an act of patriotism. As we have discussed with homonormativity, white, affluent, gay, cisgender males have become the face of the queer community. Gender, race, and class intersect to target white, middle class, gay males as ideal national capitalist consumers.

Therefore, gay marriage was commodified to target white males and to further perpetuate the idea of the nuclear family. There are multiple modalities through which gay marriage has been normalized into a consumeristic institution. Gay marriage has been co-opted by corporations and is used to create a specific image within the market, projecting it as a gateway into U.S. inclusivity. By spending money on products promoting gay marriage, one becomes a participant in the advancement of sexual modernity. A prime example of this could be seen through American Apparel's "Legalize Gay" products and advertisements, such as an advertisement featuring a young, white male wearing a shirt with the wording "Gay O.K."¹⁷ The products that these heteronormative companies produce, along with the institution of marriage and its expenses, create a pathway for certain bodies to obtain agency within U.S. society. Gay marriages contribute to capitalism and have effectively halted queer activism and severely limited the understandings of queer issues. Since the legalization of gay marriage by the Supreme Court, the nation state can now claim that the fight for equality and inclusion is complete. In reality, this is not the case; violence is enacted every day against queer subjects inflicted through institutional homophobia and transphobia embedded within the justice, financial, and larger cultural systems.

Although advocating for same-sex marriage is not the only issue within queer communities, the push for marriage equality was seen as the main issue brought to mainstream society. Not all queer individuals desire to be married; however, desires for marriage equality by normative queer individuals appeal to the nation's ideals. The movement to secure same-sex marriage rights was predominantly advocated for by white, elite members of the queer community, and integrated into popular US discourse, despite being an issue with which many periphery groups within the queer community do not agree. As Chandan Reddy argues, the United States adopted the movement for marriage equality to replace previous civil rights movements in order to prove the nation's modernity. Specifically, Reddy writes of the passing of Proposition 8 in California:

This is a moment in which, as Foucault argues of modern power, the state originates from below--is appropriated by subaltern groups, no longer entirely or exclusively in subalternity, for their own interests. We are witnessing the possibilities of transnational public sources and cultural heterogeneities for remaking state power against both the neoliberal and the welfare state...

This is not about rights, equality, or identity--it is about the speech of bodily groups that are the material foundations of the US nation-state.¹⁸

In Reddy's argument, certain subjects who advocated for Proposition 8 in California were able to appeal to the neoliberal state and further the image of same sex marriage as an issue relevant to all members of the queer community. Marriage equality was not the most imperative issue to all queer individuals; however, it was adopted by the nation state and by mainstream media because it aligns with other national ideals and has the largest potential to perpetuate the modernity of the United States.

Current Acts of Queer Assimilation

In making definitive claims about the LGBT community and their budding relationship with the nation, it is necessary to locate movements within the community that not only assert normative gays as a member of the nation but also perpetuate American exceptionalism and Islamophobia, as these are markers of homonationalism. A current example of this is the amount of LGBT groups in support of Republican President-elect Donald Trump. These groups became quite vocal in the support of Trump after the attacks on Pulse night club in Orlando, as many of them claimed that Trump would be the best option to protect them for the threat of ISIS. As the former president of a gay conservative group, GOProud, Chris Barron claimed two days after the shooting: "I have no doubt that Donald Trump would be better for LGBT Americans, when we've got ISIS throwing gay people off of buildings, when we have Muslim states that are prescribing the death penalty for people who are gay, I would think this would be something that a friend of the LGBT community would be able to speak out on."¹⁹ This statement illustrates the ways in which the nation and newly assimilated queer individuals use contrasting images against other nations to promote Western modernity. The concerns Barron mentions allow the United States to divert views of queer oppression in the nation, as Barron paints Muslim states as the most harmful force to queer bodies and the US nation as their Protectorate. Similar groups such as

LGBTrump-Gays for Trump have issued very similar homonationalist statements as seen in their bio: "As Americans from all walks of life listen to Trump's message of economic nationalism, American exceptionalism, and limited government, those of us in the LGBT community should start paying close attention," as well as their official statement on the night of Trump's victory, "Trump's victory tonight will usher in a new era of politics – an era in which your identity does not determine your ideology and an era in which the needs of the nation are placed before the needs of special interests."²⁰ These statements are directly calling for queer individuals to move their concerns beyond their queer identities in return for a national identity, as that is what would be most effective for the queer community. With said national identity these groups are calling for queer individuals to focus on issues threatening the nation rather than how the nation treats its queer individuals as the nation and its queers are one in the same now. Both Barron and gays for trump can serve as representations of how the new gay, which relies on assimilation, nationalism, islamophobia and heteronormativity is crafted in queer communities.

Conclusion

In the years following 9/11, brown and Muslim bodies have become the target of collective racism and exclusion from the nation, while the state simultaneously adopts a guise of modernity through its inclusion of an idyllic homonormative citizen. This reorganization of citizenship allows the nation to "craft the new gay" and to promote US exceptionalism. This is accomplished by the state's endorsement of feelings of anxiety by its "neurotic citizens," while failing to recognize the emotions of and marginalization of non-white bodies. The inclusion and participation of certain gay subjectivities within the nation, dubbed as homonationalism by Puar, is achieved through the exclusion of marginalized groups from what constitutes an idealized and desirable neoliberal citizen.²¹ Because of the link between capitalism and a neoliberal citizenry, print ads and commercials spell out and take on the nationalist agenda of "crafting the new gay" through normative representations of select queer subjects. In addition, the co-optation of issues, specifically the gay marriage movement, not only limits an understanding of what queer issues are, but also encourages the inclusion of homonormative subjects into the consumerist institution of marriage, providing an opportunity for ideal gay subjects to participate and further an exclusionary nationalist agenda. By crafting

this new gay citizen, the United States is able to position itself as culturally superior under the guise of sexual modernity.

References

- 1Najmabadi, A. (1998) "Crafting an Educated Housewife in Iran." In *Making Women: Feminism and Modernity in Middle East*. Abu-Lughod, L. (ed). Princeton University Press. Pg 91-125.
- 2Puar, J. (2006) "Mapping US Homonormativities." *Gender, Place, and Culture*. 13.1. Pg 67-88.
- 3Puar, J. (2013) "Rethinking Homonationalism." *International Journal of Middle East Studies*. 45. Pg. 336-339.
- 4Johnson, C. (2010) "The Politics of Affective Citizenship: From Blair to Obama." *Citizenship Studies*. 14.5. Pg 495-509.
- 5Ahmed, S. (2011) "Problematic Proximities: Or Why Critiques of Gay Imperialism Matter." *Feminist Legal Studies*. 19. Pg 119-132.
- 6Ahluwalia, S. (2015) "Abortion and Gay Marriage: Sexual Modernity and its Dissonance in Contemporary World." *Economic and Political Weekly*. L.50. Pg 27-30.
- 7Puar, J. (2004) "Abu Ghraib: Arguing Against Exceptionalism." *Feminist Studies*. 30.2. Pg 522-534.
- 8Ahluwalia, S. (2015) "Abortion and Gay Marriage: Sexual Modernity and its Dissonance in Contemporary World." *Economic and Political Weekly*. L.50. Pg 27-30.
- 9
- 10Duggan, L. (2002) "The New Homonormativity: The Sexual Politics of Neoliberalism." In *Materializing Democracy: Toward a Revitalized Cultural Politics*. Castronovo R., and Nelson, D., eds. Duke University Press. Pg 175-194.
- 11 Stokes, Zeke. Where We Are On TV. GLADD. 2015. Accessed November 17. <http://www.glaad.org/files/GLAAD-2015-WWAT.pdf>
- 12Kidd, D. (2014) *Pop Culture Freaks Identity, Mass Media, and Society*. Westview Press. Pg 129-165.
- 13 Gay Press Report. Riverdell. 2016. Accessed November 17. <http://rivendellmedia.com/wp-content/uploads/2016/10/Gay-Press-Report-2015-Condensed.pdf>
- 14N.A. (2013) "Kindle Gay Amazon Commercial – Out Now Business Class – Better LGBT." *Youtube*. Amazon. <https://www.youtube.com/watch?v=wY1UIES9wx8>. (Accessed 04/25/2016).
- 15 N.A. (2010). "Great to See You Out." Miller Lite. SAB Miller. <https://openlygaymarketing.files.wordpress.com/2010/09/millerlitecouple.jpg?w=240>. (Accessed 04/25/2016).
- 16N.A. (2010) "Second Annual GLAAD Media Awards in Advertising Nominees." GLAAD.ORG. GLAAD. <http://www.glaad.org/advertisingawards/2010/nominees>. (Accessed 04/25/2016).
- 17N.A. (2016) "Legalize Gay." American Apparel. http://store.americanapparel.net/en/legalize-gay_legalize-gay;jsessionid=589EAE9293425C11916603A6A6DD3038.aap-prd-dal-app-01-p-app1. (Accessed 4/25/2016).
- 18 Reddy, C. (2011) "Moving Beyond a Freedom with Violence: The Politics of Gay Marriage in the Era of Radical Transformation." *Freedom with Violence: Race, Sexuality, and the US State*. Duke University Press. Pg. 182-218.
- 19 Moody, Chris interview with Chris Barron. CNN. June 15, 2016
- 20LGBTrump- gays for trump's Facebook Page, Accessed November 17, 2016, https://www.facebook.com/LGBTrump/?ref=page_internal
- 21Puar, J. (2006) "Mapping US Homonormativities." *Gender, Place, and Culture*. 13.1. Pg 67-88.

Conceptualizing Trauma by Embracing Ambiguity

BY AUDRA DEBOY, BA
GETTYSBURG COLLEGE

Abstract

Debate over Posttraumatic Stress Disorder (PTSD) has intensified as experts voice persistent disagreement over the conceptualization, etiology, and treatment of PTSD. Ruth Leys, author of Trauma: A Genealogy, argues that no resolution to the debate over how to conceptualize trauma is likely to emerge since theorists inevitably oscillate between psychological and neurobiological theories. This ambivalence, she claims, indicates that therapeutic practice should not follow theory but instead, should ignore the relation between theory and practice and explore a variety of treatments. Chris Brewin and Matthew Friedman, two experts involved in the development of PTSD in the DSM-V, each contribute to the theoretical discussion about PTSD by their two distinctive approaches. In examining such theoretical differences, I argue that the debate surrounding trauma is part of the productive process in the creation of one of its many different conceptualizations, mediated by socio-cultural, historical, and scientific contexts. Such theoretical tension is valuable in suggesting very different models for conceptualizing PTSD influences and alleviating debilitating symptoms in suffering individuals.

Introduction

How psychological trauma is conceptualized determines who is considered to be traumatized, and whether or not—as well as how—they are treated. Changes in definition and criteria, as laid out in the diagnosis for PTSD within the *Diagnostic and Statistical Manual of Mental Disorders* (DSM), directly impact the rates of prevalence, as some who were previously identified with the disorder no longer meet the criteria for PTSD, though they may share similar symptoms. In a recent national study on prevalence rates of PTSD diagnoses due to the shift in criteria from the DSM-IV to the DSM-V, researchers found that there was a decrease in prevalence rates due to this shift in criteria.¹ This study illuminates the fragility of the concept of trauma and its importance as defined by the DSM, since it directly impacts those individuals who either fit or do not fit the diagnostic criteria. Psychologists and physicians alike grappled with a way to conceptualize trauma as future researchers and clinical practitioners continue to struggle to create more adequate conceptualizations of trauma-related disorders.

In what follows, I will trace the evolution of trauma, present current theories on how it should be conceptualized, and contend with a skeptical view on the depth of theoretical disagreement. The first section describes the evolution of the concept of trauma, beginning in the mid-19th century with John Eric Erichsen's description of the effects of a series of railway accidents and ending with today's current conceptualization as described in the DSM-V.² Next, I expound Ruth Leys' skeptical argument which states that because there is an oscillation between psychological and positivist theories of trauma, hope of ever developing one conceptualization is bound to fail.³

Furthermore, she argues that practice should not follow theory since conceptualizations of trauma oscillate between these two theories. I then juxtapose two current experts on the concept of trauma, Chris Brewin and Matthew Friedman, with Leys' theory. Though both Brewin and Friedman agree that there is likely to be a more adequate conceptualization of trauma which may be identified with more research, they disagree with one another on how best to currently conceptualize trauma. Chris Brewin takes a narrow approach to trauma, arguing that there should be no stressor criterion and fewer symptoms for the diagnostic criteria of PTSD, whereas Matthew Friedman takes a broad approach, arguing that there should be a specific set of stressors and a large list of symptoms. I argue that debate surrounding trauma is an essential part of the productive process in the creation of one of its many accurate conceptualizations as mediated by its socio-cultural and historical context, which will influence treatment methods that are functionally useful in alleviating suffering individuals of their debilitating symptoms.

History of Trauma

The term *trauma* from the Greek word *τραῦμα* originally referred to a physical wound, though its meaning has now evolved to represent a psychological wound as well. However, the metaphorical use of the term *trauma* complicates efforts to theorize trauma. From its inception, psychological trauma has been couched in terms of physical wounds before it itself was fully fleshed out and its underpinnings understood. John Eric Erichsen, a mid-19th century British surgeon, described the set of symptoms occurring in victims of a series of railway accidents in physiological terms, describing trauma

as a “concussion of the spine.”⁴ While it is possible that neurobiologists will reduce the symptoms which a traumatized individual experiences to a neurobiological event, this is not yet certain.

The initial conceptualization was further complicated when psychologists and psychoanalysts began describing symptoms in terms of changes within the psyche. One British military psychiatrist, Charles Samuel Myers, rejected the notion that the experience of trauma resulted from an organic cause.⁵ Rather, he approached the conceptualization of trauma from a psychological standpoint, using the term, “shell shock,” to refer to the severe emotional effects that trauma causes. Thus, Myers played an important transitional role in connecting the older, organic theories of origination to the later psychoanalytic and psychological theories of causality. The late-19th century neurologist Jean-Martin Charcot began conceptualizing trauma by borrowing hysteria from its religious context and importing it into a scientific one.⁶ With his term “hystero-traumatic auto-suggestion,” Charcot proposed that traumatic events caused traumatized individuals to be in a suggestible, hypnotic state which also resembled the dissociative aspects of hysterical symptoms.⁷ Though some of Charcot's students at the Pitié-Salpêtrière Hospital in Paris focused on suggestibility and simulation in relation to trauma, Sigmund Freud studied and incorporated childhood sexual abuse and hysteria into his psychological account of trauma.

Sigmund Freud further developed his teacher's theory, locating the emergence of the hysterical symptoms within the individual's psyche rather than deriving from the traumatic experience itself. From a case study of eighteen individuals with symptoms

of hysteria, Freud concluded that symptoms of hysteria arise out of the individual who represses memories of early childhood sexual abuse or experiences.⁸ Experience, according to Freud, causes the repressed feelings and symptoms from an earlier childhood trauma to surface.

Despite efforts to give a psychological account of traumatic symptoms, many did not accept these symptoms as real or valid. It was not until after the Vietnam War that a third edition of the DSM was established, which made Posttraumatic Stress Disorder an official term after immense pressure from psychiatrists and activists.⁹ Two such activists, Ann Burgess and Lynda Holstrom, completed work on sexual and domestic violence, raising awareness about these types of trauma to the public.¹⁰ Also studying the effects of sexual abuse on children, Judith Herman drew similarities between the symptoms of children who had been sexually abused and veterans who experienced traumatic events.¹¹ Additionally, Herman advocated for Three-Phase Trauma Therapy, an individualized treatment option for trauma survivors.¹² These feminists, among other activists, made the public and the scientific community aware of pervasive child sexual abuse and domestic violence issues, providing evidence for the similarities between the effects of these abuses and the effects of physical violence on soldiers due to the war. As a result of their efforts, PTSD became a diagnosis within the DSM-III, solidifying emphasis on psychological aspects of trauma and trauma-related injuries, as well as affirming the detrimental effects that war had on soldiers. Thus, the conception that soldiers who returned from war were predisposed to mental illness was challenged by a new account of the origin of PTSD.

The most recent diagnostic manual, the DSM-V, provides in-depth, diagnostic criteria for PTSD, as well as a number of other trauma- and stressor-related disorders. In order to qualify for a diagnosis of PTSD, an individual must first meet the criteria of having experienced “actual or threatened death, serious injury, or sexual violence.”¹³ Thus, the conceptualization of trauma has evolved within a psychological framework, with researchers working hard to establish a neurobiological foundation for PTSD symptoms.

Although the body can be observed directly, we can only observe the mind indirectly; it resists easy observation. Because of this fact, it is difficult to determine which, if any, fundamental theory of trauma captures its reality. Moreover, the tendency to reduce PTSD to psychological or neurobiological conceptualizations may limit the way in which trauma—both its

causes and the subsequent psychological effects—is understood; limited theories may neglect or fail to recognize some aspect of trauma (and its effect on the individual) that is an integral part of the experience of trauma. The issue is further complicated by the relation between the traumatic event and its effect on the individual. Since not all individuals exposed to the same traumatic event will become traumatized, it is difficult to determine the degree to which trauma is due to the stressor itself, the susceptibility of the traumatized person, or other factors related to social context. Thus, trauma’s conceptualization is subject to great debate, since the mind is not fully understood nor directly observable, and since the relation between stressor and traumatized individual is subject to social and psychological factors.

Current Conceptualizations of Trauma

Tracing the evolution of conceptualizations of trauma reveals an ongoing struggle to identify the underlying reality of trauma. Starting as a physical wound, trauma as a concept has transformed to represent a wound of the individual and collective psyche as well. Still today this is a hotly contested concept, with conflicting views about its core symptoms, whether or not it should be considered its own diagnosis in relation to other disorders, and what types of stressors that should count toward the diagnosis of PTSD.

Ruth Leys, author of *Trauma: A Genealogy*, approaches the concept of trauma through a genealogical lens in order to elucidate its long and interesting history. She claims that the debate is stagnant since experts repeatedly divide between those who defend what she calls the *mimetic* theory and those who defend the *anti-mimetic* theory.¹⁴ She uses the term *mimetic* to refer to psychological theories of trauma and the term *anti-mimetic* to refer to neurobiological theories of trauma. Hence, she claims “discussions of trauma are characteristically polarized between competing positions each of which can be maintained in its exclusiveness only at the price of falling into contradiction or incoherence.”¹⁵ From this argument, she makes a claim that therapists should not exclusively rely on treatment techniques which derive from current Western theories of trauma since the theories themselves are inconsistent.¹⁶ Rather, Leys advises a pragmatic approach to treatment that allows for experimentation in practice, calling for the use of a variety of treatment techniques.

Debates among contemporary experts on PTSD reflect similarly complex issues about PTSD. On one side of the controversy about the conceptualizations of trauma, Chris Brewin, professor of Clinical Psychology at the University College London, takes a narrow

approach, as evidenced by diagnostic criteria he proposes for PTSD. In contrast, Matthew Friedman, professor of both Psychiatry and of Pharmacology and Toxicology as well as executive director of the U.S. Department of Veterans Affairs National Center for Post-Traumatic Stress Disorder, defends a broader definition of trauma. If we situate Brewin and Friedman’s methodological approaches in relation to Leys’ theory, their positions reiterate the tension in previous theories of trauma and do little to move the debate closer to resolution. Three of the most controversial points within the contemporary debate on trauma that Brewin and Friedman represent are whether or not PTSD should be considered its own diagnosis, which symptoms should be included within this diagnostic criterion, and which stressors should count as contributors to the subject’s experience of trauma within the PTSD diagnostic criteria.

Brewin argues for a narrow construct of PTSD with the elimination of Criterion A1, which specifies the stressor criteria as well as simplifying the criteria’s set of symptoms in order to distinguish PTSD from other mental disorders. Both conceptually and clinically, a core set of symptoms for the PTSD diagnostic criteria would be advantageous because “the removal of symptoms associated with general dysphoria should also lead to greater homogeneity of cases and reduced overlap with other disorders.”¹⁷ Many of the symptoms listed within the criteria for PTSD are shared with a number of other disorders, such as the anxiety disorders. Thus, Brewin argues for a simplification of the current diagnostic criteria for PTSD in order to capture the defining features of PTSD and increase the reliability with which it is diagnosed. A refocusing of the symptom criteria would “highlight the features that are most salient to the individual with PTSD, that are the primary focus of psychological treatment.”¹⁸ The nature of trauma is such that, he argues, it should be considered a distinct disorder, since it has a pathology which arises out of the persistence and impairing effects of its distinct set of symptoms and activation patterns within the brain.

In contrast to Brewin’s narrow definition of trauma, Friedman offers a broad conceptualization of trauma, which expands the symptom criteria for PTSD within the DSM-V. Rather than eliminate symptoms which overlap with other disorders, Friedman says they should remain within the diagnostic criteria for PTSD, claiming, “if one extended this logic to medical diagnosis, one would eliminate symptoms such as fever, pain, and edema from the diagnostic criteria of a specific disease because they are found

in so many other diseases.”¹⁹ It would not make sense to eliminate symptoms which accompany PTSD on the grounds that they are also exhibited in other disorders; distinct symptoms and unique combinations of symptoms will distinguish PTSD from other mental disorders. By maintaining a broad PTSD construct, therapists are able to diagnose based on an array of symptoms, including those symptoms which are important to PTSD despite being shared with other disorders. In conclusion, Friedman asserts a broad conceptualization of trauma and preserves the stressor criteria within the diagnostic criteria for PTSD as an integral part of the diagnosis.

The 11th edition of the International Classification of Diseases (ICD-11), created by the World Health Organization, and the DSM-V mirror Brewin and Friedman’s debate between a narrow and broad conceptualization of trauma. While the DSM-V uses the broad construct of trauma for PTSD as described by Friedman, the ICD-11 is more closely related to Brewin’s narrow construction. Like Brewin’s proposal, the ICD-11 has eliminated general symptoms shared with other disorders and narrowed down the symptom list to three symptom clusters.²⁰ These clusters include re-experiencing the trauma, aversion to reminders of the trauma, and a perceived current threat resulting in hypervigilance.²¹ Additionally, one element not seen in the DSM-V that will be in the ICD-11 is complex PTSD. This diagnosis is for those individuals who meet the regular criteria for PTSD but who also have additional symptoms, such as somatization, dissociation, shame, guilt, and interpersonal problems.²² Even regions of the world are at odds with the conceptualization of trauma, with the broad construct more common in the U.S. and the narrow construct more common internationally.

The historical development of the concept of trauma, with its persistent tensions, has given rise to the current psychological, psychiatric and neurobiological irresolution in debates over the DSM-V. Some experts are discouraged by the inability to arrive at a stable concept of trauma, whereas others maintain hope that with more research, a settled concept of trauma will emerge. Leys belongs to the former class, arguing that conceptualizations of trauma are likely to continue to oscillate between *mimetic*, or psychological and hypnotic-state, and *anti-mimetic*, or scientific and positivist, theories. In contrast to the skeptical view put forth by Leys, Brewin and Friedman, though they disagree on the conceptualization of trauma, agree that by continued exploration researchers will gain a better understanding of trauma. The current psychiatric conceptualization

of trauma, understood by specific criteria for PTSD, is currently under scrutiny for bracket creep, the excessive breadth of its conceptualization and what counts as a traumatic event. Brewin argues that the stressor criterion should be eliminated in favor of a core set of symptoms that constitute the criteria of PTSD. In contrast, Friedman defends a broad conceptualization that encompasses the entire set of symptoms relevant to the posttraumatic phenotype.

An Alternative Approach to Trauma

Experts on trauma have been unable to agree on a working concept of trauma that guides clinical practice. Theoretical disagreement surrounding the conceptualization of trauma has tended either toward a psychological approach or a neurobiological understanding. Although Leys is skeptical about the possibility of a resolution, contemporary experts like Brewin and Friedman, who are working on the diagnostic criteria for PTSD, maintain hope that close analysis will lead them toward a clearer understanding of psychological trauma and its causes. Additionally, Leys argues that clinical treatment practices for PTSD should not follow theory, freeing clinical practitioners from the impasse of theory. Rather than assuming a persistent and irresolvable tension in conceptualizations of trauma, deepened by the historical and scientific contexts out of which they emerge, I believe a more productive approach would be to welcome a wide range of treatments so that greater evidence of efficacy can be established.

It is too hasty to dismiss theories of trauma on the basis that disagreement exists among experts, as Leys does. If we were to abandon any concept which does not present itself in a clear manner from its inception, there would be no room for intellectual growth. Mark Silcox writes:

If biologists were entitled to talk about “genes” long before the structure of DNA was discovered, and if contemporary physicists can afford to feel comfortable extending the concept of a “particle” to units of gravitational force, it is difficult to see why proleptic speculation about consciousness-producing neurons or ‘trauma in the soul’ should be regarded as any more problematic.²³

Just as the concept of genes was posited in the scientific vocabulary as an educated assumption about the conceptual schema prior to the discovery of DNA, conceptualizations of trauma have a valuable role, however problematic they may be. Though theoretical disagreements persist, we should remain optimistic that with more research, the precipitating factors and psychological

symptoms of trauma will be understood better. In the meantime, we should also be receptive to evidence regarding the functional utility of current concepts of trauma that prevail in clinical practice.

What Leys views as a conceptual impasse might therefore better be seen as part of the creative process in trauma’s conceptualization. In its current state in the DSM-V, trauma in terms of PTSD criteria is highly contested for its stressor criterion, as illustrated in Brewin and Friedman’s earlier debate. Since current methodologies are incapable of detecting observable changes in an individual’s psyche that has been affected by a traumatic event, it is advantageous in practice to maintain multiple theories of trauma. Too much ignorance surrounds the concept of trauma for any expert to pin down one specific theory. Instead, researchers need to consider all of the conceptualizations which can be justified using available evidence. Moreover, the debate around trauma reflects the potential directions future research may take. Despite referencing the same evidence, experts often disagree, as do Brewin and Friedman. When this happens, Hilary Kornblith, professor of Philosophy at the University of Massachusetts Amherst, says, “one ought to suspend judgment.”²⁴ It is not that the debate on trauma has been polarized between psychological and neurobiological approaches, but rather that experts have not yet agreed because evidence has not yet settled unresolved issues. In trauma’s current conceptualization there is understandable disagreement since there is still much to learn about what gives rise to trauma, what kinds of events should be described as trauma inducing, and how we should think about the role played by the psychology of the individual.

What appears to be an impasse for Leys is actually productive for both the neurobiological and psychiatric communities that assert their individual claims about the nature of trauma. Psychologists and psychiatrists attempt to explain the traumatized individual’s experience in terms of mental phenomena.²⁵ However, because mental phenomena cannot be directly observed, psychologists have to rely upon personal accounts of the traumatized individual’s experience and summarize this in figurative language. In contrast, neurobiological accounts favor brain-based explanations for the symptoms of PTSD, such as van der Kolk’s biological theory of traumatic memory.²⁶ Despite the contrary views on trauma, this is not to say that one view is superior to the other or that progress is at a standstill because there are multiple views. As Silcox observes, “the fact that scientific psychology and our everyday

mentalist idioms are still haunted by the specter of Descartes should surely not prevent clinicians from extending into the sciences of the mind the application of terms already used to classify more naturalistically tractable phenomena.²⁷ Thus, Silcox argues that despite the problem of the mind-body dichotomy which still plagues much of the scientific field, researchers should deconstruct this tired idea and take epistemic leaps to assume a neurobiological basis for some of the psychological phenomena, just as biologists did with genes before DNA was understood. Efforts such as these will draw new connections between mind and body conceptualizations of trauma and give rise to yet another perspective on trauma, approaching it this time from an intermediary position.

Leys' expectation that trauma should have a settled conceptualization is further complicated by the evolving contexts out of which conceptualizations of trauma emerge. There may never be one concrete set of diagnostic criteria for PTSD. Instead, criteria may need to be adjusted as both practice and theory evolve in different socio-cultural contexts. The stressor criterion debate revolves around whether or not there should be a criterion which specifies which events count as contributors to the experience of trauma for the individual who experiences it. The latest edition of the DSM does include a stressor criterion which specifies what events count as "traumatic." It is easy to imagine that with a concept like trauma, its conceptualization will be fluid, evolving out of the socio-cultural and historical context in which it develops. To elaborate, the DSM did not include PTSD until psychiatrists and other individuals advocated for a separate category which accounted for the suffering of individuals returning from the Vietnam War.²⁸ It is clear that context influenced trauma's psychiatric conceptualization both in terms of how researchers conceptualized it, as well as in terms of who could be considered traumatized under this new diagnosis. According to Allen Young, PTSD is "glued together by the practices, technologies, and narratives with which it is diagnosed, studied, treated, and represented and by the various interests, institutions, and moral arguments that mobilized these efforts and resources."²⁹ For example, combat soldiers returning from war were some of the first individuals to be considered as having experienced a traumatic event with the potential to develop PTSD. However, as technology changes and soldiers increasingly end up behind a screen operating remotely piloted combat drones, these individuals are expressing PTSD symptoms and high levels of stress similar to those of combat soldiers.³⁰ Thus, as new technologies are

developed and new instances arise in which an individual may become traumatized, the conceptualization of trauma must be adjusted to take account of these cases.

Despite theoretical disagreement, treatment options that follow theory should be considered for individuals with PTSD. Leys is too quick to dismiss these treatment options, as she claims:

If it is true that the entire discourse on trauma in the West has been structured by an unresolvable tension or conflict between mimesis and antimimesis, then it would be a mistake for therapists to think that treatment for the victims of trauma should follow theory in some direct way, because that theory will continue to be subject to the alternations and contradictions inherent in the *mimetic-antimimetic* structure.³¹

Though safe, alternative therapies should be considered for individuals with PTSD; Leys fails to recognize that traumatized individuals cannot wait for theory to catch up to practice. Regardless of whether or not any particular theory is confirmed, treatments following theory have shown effectiveness in many cases in relieving traumatized individuals of their symptoms. One meta-analysis on the efficacy of psychological treatments for PTSD found that many current treatment methods including, "exposure-based therapy, CPT [cognitive processing therapy], cognitive restructuring, CBT-mixed therapies [cognitive behavioral therapy], NET [narrative exposure therapy], and EMDR [eye movement desensitization and reprocessing]" resulted in large reductions in PTSD symptoms.³² This is not to say that because some treatment methods do relieve the symptoms of some individuals with PTSD, efforts to conceptualize trauma are unnecessary or that they should stop. Furthermore, though current treatment methods that follow theory already prove effective for many traumatized individuals, more accurate and effective treatment methods may be developed. As researchers uncover the underpinnings of trauma, priority should be focused on suffering individuals and treatments that prove safe and effective in reducing symptoms of PTSD.

In addition to treatment methods which prove effective for relieving symptoms of PTSD, other safe and effective treatment methods should be considered as well. Since trauma is not yet well-understood, it is reasonable to take a variety of approaches with the hope of learning more about it. This includes challenging the intellectual convention for conceptualizing mental disorders and exploring treatment options outside of psychology and psychiatry. One such alternative treatment option for PTSD is Narrative Adaptation. MaryCatherine

McDonald, visiting professor of Philosophy at the College of the Holy Cross, claims that "trauma disrupts the ability of the subject to synthesize, categorize, and attach meaning to events," and that narrative enables the traumatized individual to record, synthesize, and assign personal meaning to the event, breaking the reappearance of the event as a present happening.³³ McDonald claims that through narrative, or the telling and retelling of the event, the traumatized individual is able to regain agency and view the event as occurring in the past. Even if the explanation is incomplete for why narrative therapy alleviates some symptoms of PTSD, therapists have shown increasing interest in it as a promising new treatment.³⁴ In addition to having at least the ability to ameliorate the lives of individuals with PTSD, alternative treatment methods such as these may also give researchers some insight into the way in which trauma affects individuals. Alternative treatment options, like Narrative Adaptation, show promise in alleviating traumatized individuals of their symptoms and should be explored further.

Another longitudinal study examined individuals who have experienced traumatic events and observed biopsychosocial effects of trauma, as well as moral and spiritual effects. Examining the effects of political violence in KwaZulu-Natal over the course of five years, the results of the Harvard Trauma Questionnaire suggested that participants experienced a moral and spiritual weakening after exposure to trauma.³⁵ Specifically, they claim that a traumatic event injures one's moral beliefs since it "creates dissonance and conflict because it violates assumptions and beliefs about right and wrong and personal goodness."³⁶ Additionally, the traumatic event can affect "the faith that God is constantly available to respond to one's hopes, fears, anxieties and tragedies... [is] shattered. When this happens, individuals who are unable to resolve challenges to their moral and spiritual beliefs might find themselves in a state of spiritual alienation."³⁷ The traumatized individual who previously had a spiritual connection to God, for instance, may feel that God has neglected them, leading to their rejection of God or system of faith. Based on these findings, the researchers argue for a treatment method which takes into account such challenges to moral and spiritual beliefs. Perhaps the individuals subject to political violence would have benefitted from a communal approach to healing similar to the method used by Urban Warriors of Chicago. Urban Warriors pairs children subject to violence in their neighborhoods with veterans who teach their students coping mechanisms to alleviate PTSD-like symptoms.³⁸ This approach to PTSD captures effects of

trauma that are overlooked by the traditional biopsychosocial approach to mental disorders. Such alternative approaches to trauma that explore a wider range of effects of trauma are invaluable in the exploration and conceptualization of trauma. Not only do such alternative approaches have promise in aiding understanding of trauma, but they suggest novel treatment methods.

Conclusion

Since its clinical inception in the mid-nineteenth century, the concept of trauma has been subject to many fluctuations, causing some experts to become skeptical that there will ever be an adequate concept of trauma. What was initially conceptualized as a concussion of the spine by John Eric Erichsen after studying a group of railway accident victims eventually took on the biopsychosocial understanding present today. Along the way, it transitioned between physiological explanations, like “traumatic neurosis” coined by Herman Oppenheim, and psychological ones such as Charles Samuel Myers’ term ‘shell shock.’³⁹ In his link of trauma to hysteria, Freud later maintained that individuals with a certain disposition would repress early childhood sexual experiences, which would distress the ego and produce symptoms of hysteria during a triggering event in puberty. Thus, conceptualizations of trauma have fluctuated between psychological and biological manifestations, often as the result of social change. For instance, during World War I, trauma was understood to be the result of a physiological change since stigma attached to psychological explanations. After the Vietnam War, PTSD was only added into the DSM once psychiatrists and other activists advocated for a separate diagnosis for individuals returning from war with similar dysfunctional symptoms. Despite skepticism by some trauma experts that researchers will be able to develop an adequate concept of trauma, other experts work to develop an adequate conceptualization of trauma. Although much theoretical disagreement exists, this should not discourage further research and analysis.

Not only have treatment options improved over time to include a more holistic approach to treating the symptomatology of PTSD and reduce the suffering of traumatized individuals, but also current treatment options have proved effective for reducing PTSD symptoms. I do not believe the concept of trauma is fated to a conceptual impasse. Rather, conceptualizations of trauma are multifaceted because psychological trauma cannot be observed like a physical wound, and our current understandings include both psychological and physical aspects. Thus, I

argue for greater theoretical breadth and exploration in approaching trauma. Over time the conceptualizations of trauma will inevitably fluctuate as theory adjusts to a changing world, as seen in accounts of trauma that result in PTSD-like symptoms in those who operate remotely piloted aircrafts. There may never be a single settled conceptualization of trauma, but rather a series of theoretical approaches to a broad range of trauma-related symptoms. Diagnostic categories may be expanded upon, reassigned, or eliminated completely, new stressors may be identified, and new effective treatments employed. Just as the term “gene” was assumed prior to a full understanding of DNA, trauma has to be described before we arrive at a more settled conceptualization of its causes. Beyond debates over theory and practice, we must remember that there are suffering individuals and seek always to respond with treatment that mitigates trauma-related symptoms. When Leys argues that practice should not follow theory, she ignores the fact that many treatments that follow theory, even if that theory may not yet capture the full reality of trauma, are functionally effective. Although current therapeutic methods should continue to be used until more effective methods are discovered, a wide variety of approaches should be researched and evaluated that reflect a more holistic view of the suffering individual. In conclusion, the debate surrounding trauma and its diagnostic criteria and best treatment reflects a productive process that gives rise to a richer and more varied conceptualization of trauma. Such richer conceptualizations carry promise for new approaches to treatment that alleviate the debilitating symptoms of those who have suffered from posttraumatic stress.

References

¹Kilpatrick, D.G., Resnick, H.S., Milanak, M.E., Miller, M.W., Keyes, K.M. and Friedman, M.J. (2013) “National Estimates of Exposure to Traumatic Events and PTSD Prevalence Using DSM-IV and DSM-5 Criteria.” *Journal of Traumatic Stress*. 26.5. Pg 537-547.
²Leys, R. (2000) *Trauma: A Genealogy*. University of Chicago Press. Pg 3.
³*Ibid.* Pg 305.
⁴*Ibid.* Pg 3.
⁵van der Kolk, B.A. (2007) “The History of Trauma in Psychiatry.” *Handbook of PTSD, First Edition: Science and Practice*. Friedman, M.J., Keane, T.M., and Resick, P.A., eds. Guilford Press. Pg 19-36.
⁶Reuther, B.T. (2012) “Philosophy of Trauma.” *In Encyclopedia of Trauma: An Interdisciplinary Guide*. Figley, C.R., ed. SAGE Publications. Pg 437-441.
⁷van der Kolk, B.A. Pg 21.
⁸Freud, S. (1962) “The Aetiology of Hysteria.” *In The Complete Psychological Works of Sigmund Freud*. Strachey, J., ed. The Hogarth Press Limited. Pg 188-221.
⁹Leys, R. Pg 5.
¹⁰Ray, S.L. (2012) “Posttraumatic Stress Disorder, History of.” *In Encyclopedia of Trauma: An Interdisciplinary Guide*. Figley, C.R., ed. SAGE Publications. Pg 457.
¹¹Leys, R. Pg 5.

¹²Reuther, B.T. (2012) “Philosophy and Ethics of Trauma Treatment.” *In Encyclopedia of Trauma: An Interdisciplinary Guide*. Figley, C.R., ed. SAGE Publications. Pg 435-436.
¹³American Psychiatric Association. (2013) *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. American Psychiatric Association. Pg 265-291.
¹⁴Leys, R. Pg 305.
¹⁵*Ibid.* Pg 305.
¹⁶*Ibid.* Pg 307.
¹⁷Brewin, C.R., Lanius, R.A., Novac, A., Schnyder, U. and Galea, S. (2009) “Reformulating PTSD for DSM-V: Life after Criterion A.” *Journal of Traumatic Stress*. 22.5. Pg 371.
¹⁸*Ibid.* Pg 366-373.
¹⁹Friedman, M.J. (2013) “Finalizing PTSD In DSM-5: Getting Here From There And Where To Go Next.” *Journal of Traumatic Stress*. 26.5. Pg 548-556.
²⁰*Ibid.* Pg 554-555.
²¹Karatzias, T., Shevlin, M., Fyvie, C., Hyland, P., Efthymiadou, E., Wilson, D., Roberts, N., Bisson, J.I., Brewin, C.R., and Cloitre, M. (2016) “An initial psychometric assessment of an ICD-11 based measure of PTSD and complex PTSD (ICD-TQ): Evidence of construct validity.” *Journal of Anxiety Disorders*. 44. Pg 73-79.
²²Valent, P. (2012) “Trauma, Definitions of.” *In Encyclopedia of Trauma: An Interdisciplinary Guide*. Figley, C.R., ed. SAGE Publications. Pg 676-679.
²³Silcox, M. (2014) “Psychological Trauma and The Simulated Self.” *Philosophy Of The Social Sciences*. 44.3. Pg 349-364.
²⁴Kornblith, H. (2010) “Belief in the Face of Controversy.” *Disagreement*. Feldman, R., and Warfield, T.A., eds. Oxford University Press. Pg 29-52.
²⁵Silcox, M. Pg 357-358.
²⁶Leys, R. Pg 229-265.
²⁷Silcox, M. Pg 358.
²⁸Young, A. (1995) *The Harmony of Illusions: Inventing Post-Traumatic Stress Disorder*. Princeton University Press. Pg 5.
²⁹*Ibid.* Pg 5.
³⁰Chappelle, W.L., McDonald, K.D., Prince, L., Goodman, T., Ray-Sannerud, B.N. and Thompson, W. (2014) “Symptoms Of Psychological Distress And Post-Traumatic Stress Disorder In United States Air Force “Drone” Operators.” *Military Medicine*. 179.8. Pg 63-70.
³¹Leys, R. Pg 307.
³²Cusack, K., Jonas, D.E., Forneris, C.A., Wines, C., Sonis, J., Middleton, J.C., Feltner, C., Brownley, K.A., Olmsted, K.R., Greenblatt, A., Weil, A. and Gaynes, B.N. (2016) “Psychological Treatments For Adults With Posttraumatic Stress Disorder: A Systematic Review And Meta-Analysis.” *Clinical Psychology Review*. 43. Pg 128-141.
³³McDonald, M. (2013) “Trauma, Embodiment, and Narrative.” *Idealistic Studies: An Interdisciplinary Journal of Philosophy*. 42.2-3. Pg 247-263.
³⁴Sloan, D., Sawyer, A.T., Lowmaster, S.E., Wernick, J., and Marx, B.P. (2015) “Efficacy Of Narrative Writing As An Intervention For PTSD: Does The Evidence Support Its Use?.” *Journal of Contemporary Psychotherapy*. 45.4. Pg 215-225.
³⁵Manda, C. (2015) “Re-authoring life narratives of trauma survivors: Spiritual perspective.” *HTS Theologesic Studies/Theological Studies*. 71.2. Pg 1-8.
³⁶*Ibid.* Pg 5.
³⁷*Ibid.* Pg 6.
³⁸N/A. (2016) “Urban Warriors.” *YMCA of Metro Chicago*. The YMCA.
³⁹van der Kolk, B.A. Pg 19-21.

Domestic Violence and Women's Empowerment in Nepal

BY ZHICHENG HAN

BRIGHAM YOUNG UNIVERSITY

Introduction

Domestic violence

Domestic violence or intimate partner violence (IPV) is prevalent, especially in less-developed countries. IPV is defined here as male-to-female partner violence, which includes physical violence and sexual coercion; this narrower definition is necessary because data on psychological and other forms of violence were not collected in this dataset. In Nepal, over 20% of women in the surveyed sample reported experiencing physical violence, and over 90% of that physical violence was committed by a current or former husband.¹ About 15% of physical violence experienced by women in the past twelve months resulted in deep wounds, broken bones, broken teeth, or other serious injuries.¹ Pregnant women in Nepal who experience violence are less likely to seek reproductive health services, such as antenatal care, which puts them at risk of perinatal and neonatal mortality. Research shows that women's empowerment is a crucial factor in reducing domestic violence; women's empowerment, especially on a household level, is hypothesized by Sarkar to be a predictor of domestic violence, independent of other social and economic factors.² Additionally, social factors are hypothesized to be correlated with domestic violence independent of demographic variables.

Although incidents of domestic violence are rarely perpetrated outside of private settings, even the most private affairs in one's life are social. Behaviors of husbands and wives alike are influenced by social ties and family dynamics. One study showed that social ties can act as a buffer against IPV, as women who visit their families are less likely to experience domestic violence.³ In this same study, violence was seen as a last resort in husbands' attempt to control their wives' deviation from established social roles in many societies. One of the driving forces of this deviation is women's empowerment. Family or individual level variables, like education level, age at first cohabitation, and income are only a part of the variables affecting domestic violence according to Carlson's 1984 paper on the ecology of domestic violence.⁴ They affect the odds of domestic violence along with social-structural factors that include larger cultural framework and power dynamic outside of the family unit.⁴

Empowerment

The definition and measurement of women's empowerment has been evolving over time; recent studies suggest that women's ability to make decisions which affect themselves and their families is an adequate way of measuring this concept.⁵ This definition of empowerment focuses on a shift of power from the powerful to the relatively powerless. Therefore, empowerment should not be measured by health and development outcomes, as they do not necessarily indicate a change in power structure within a family. Studies show that education and participation in paid labor do not significantly alter the power structure within the family.⁶ Malhotra and colleagues propose that instead of broad-based questions on education, employment, and health decisions, studies should focus on the power dynamic within the family structure, namely on who makes decisions in a household.⁵ Therefore, in the current study, empowerment is measured with specific decision-making variables which indicate how often women make decisions in the household collected through survey methods.

Women's empowerment is associated with many positive outcomes.⁷ Empowering women helps to reduce inequalities between men and women in many developing countries. Several countries, including Nepal, have initiated programs promoting women's empowerment. Empowerment programs in Bangladesh have reduced verbal and physical abuse and violence towards women.⁸ A similar study in 2003 found that female empowerment within the community often functions independently of socio-economic conditions or women's empowerment on individual levels.⁹ Research in Bangladesh shows that increased status of women at the individual level can elevate the risk of domestic violence, while community-level female status significantly lowers the risk of violence.⁹

Social factors in Nepal

Nepal is located in the foothills of the Himalayas and consists of three ecological zones: mountain, hill, and terai (or plains). Each region has a drastically different landscape, ethnic composition, religion and language. The majority of the population is Hindu, and most of them belong to a certain

caste. Although ethnic and caste affiliations are discouraged and outright discriminations are banned, ethnic boundaries are still significant.¹⁰ Differences in caste positions affect the life experience of individuals and families. About ten percent of the population in Nepal practice a minority religion: studies show non-Muslim minority religions in Bangladesh, where ninety percent of the population are Muslim, are aligned with lower rates of domestic violence.¹¹ Urban or rural residence also plays a role in shaping community, especially in Nepal where rural communities tend to be isolated villages in rugged mountain terrain.

Community variables used in the current study include ecological region, urban or rural residence, caste position, and religion, as these are the variables available in the survey on which this paper is based. Although this collection of variables is lacking in anthropological nuance, it can paint a brief picture of the lives of these individuals, and can measure—to a reasonable extent—the community and cultural context in which one lives. In the specific context of Nepal, these factors can reasonably describe one's cultural setting.

Other demographic and social economic factors are included and controlled for in statistical analysis. Many of these are considered risk factors by previous studies: a study in Cambodia found a negative association between household standard of living and physical domestic violence. The same research also showed that women with large numbers of children and 8-13 fewer years of education than their husbands are more likely to experience domestic violence. Yount hypothesized that children are stressors who elevate the risk of domestic violence.³ In a study in India, education, age of the women, and urban residence were all associated with domestic violence. The study hypothesized that young women's behavior needed to be "checked" as they learned to behave according to the social or family norms.¹² Other factors have a more ambiguous effect: one study showed that the effect of wealth in the forms of dowry and a woman's income was complicated and often contradictory in rural Bangladesh, but education and an older age at marriage reduced the risk of domestic violence.¹¹

Women's employment was also ambiguous in its effect, because even though it increased family income, female participation in labor force might hinder their fulfillment of traditional family roles.

Based on this review of the literature, women's empowerment is modeled at the individual level to see if it is associated with domestic violence. In addition, the effects of ecological region, religion, caste position, and urban residence on odds of domestic violence are examined independently. Then, interactions between factors that impact odds of domestic violence and women's autonomy were examined. Other factors that affect the occurrence and reporting of domestic violence are: wealth, level of education, employment, women's age, age at first marriage, and number of children. It is expected that the more a culture empowers women, the less likely those women are to experience IPV; Likewise, women in regions or cultures which do not empower women might have increased odds of experiencing domestic violence.

Method

Participants

The data for this study come from the Demographic and Health Surveys (DHS) conducted in Nepal in 2011. The sample design uses a multistage cluster sample based upon an updated version of the 2001 census. The sample was designed to provide estimates of most key variables for the 13 eco-development regions. The sample includes 12,918 women between ages 15-49, who each had an individual interview. Interviews were completed for 12,674 women, resulting in a response rate of 98 percent.¹ One in three women surveyed were given the domestic violence module, yielding a sample size of 3,343. In terms of residence, 17% of the respondents live in the mountain region, 40% in the hill region, and 43% in terai. Only 26% of the residents reside in urban areas. The majority (86%) of the respondents are Hindu. The second largest religious group is Buddhist (8%). The two religions consist of the majority of Nepalese religion, sharing many practices and theologies. Muslim and Christian, (2% respectively), are major minority religions on the relative fringe of society. Most of the respondents report being part of the higher caste. The average age of women surveyed is 31, and they have 3.5 years of formal education, on average. Their average age at first cohabitation is 17. Most of the woman report they are currently working (67%), and they have, on average, 2.38 living children.

Research design

Women were interviewed in their local language with a translator present to mediate. The women were asked if they had ever experienced minor physical violence, severe physical violence, or sexual violence from their partner. These results were aggregated into a single variable where woman experiencing none of these kinds of violence are coded 0 while women experiencing any of those are coded 1.

Women's empowerment was measured by decision-making power in the household. Women were asked who makes decisions in four aspects of household decision making: (1) health care, (2) household purchases, (3) visiting relatives or friends, and (4) the husband's earnings. Each of these measures was recoded into a binary variable where women who reported having no power in making a particular type of decision were coded 0 and women who had power were coded 1. These were then summed up into a four-point scale ranging from 0 (no say in any of these matters) to 4 (have a say in all four matters).

Environmental or community factors include region, religion, ethnicity, and residency. Region is the ecological region in which the respondents live, and was measured by dummy variables in the following categories: 'mountain', 'hill', and 'terai', 'hill' being the reference category. Religion was measured in dummy variables in the following categories: 'Hindu', 'Buddhist', 'Kirat', 'Muslim', and 'Christian', 'Christian' being the reference category. Ethnicity was recorded into a dichotomous variable where privileged castes were coded 1 and middle and lower castes were coded 0. Residency was measured in a dichotomous variable where urban residency was coded 1 and rural 0.

Other control variables included education, wealth, work status, age, age at first cohabitation, and number of children. Education, age, and age at first cohabitation were measured in years. Number of children was measured with a continuous variable. Work status was coded into a dichotomous variable where currently working was coded 1 and 0 otherwise. Wealth was measured with wealth index score that gave households scores based upon their wealth compared to others in the nation. National average was coded 0, and above or below average were coded as positive or negative numbers ranging from -0.15 to 2.5.

Estimation procedure

Prior to presenting multivariate results, descriptive statistics and bivariate analyses were presented, which found significant relationships between the dependent and main independent variables. Because the

dependent variable—which was experiencing physical or sexual violence from a partner—was measured as a dichotomous variable, logistic regression was used to estimate the models. The results are indicated with odds ratios, which represent the increase or decrease in the odds of women experiencing domestic violence associated with a unit change in each category of the independent variables when holding all other variables constant. The first model shows the effect of women's empowerment on domestic violence. Next, community variables are included to determine if they influence the likelihood of domestic violence. Last, demographic factors are controlled for. After the logistic regression analysis, interactions were run between the empowerment index and ethnicity, the empowerment index and urban residency, and the empowerment index and region. This was done to test if the effect of empowerment varies across these groups, and to see if individual empowerment and community level empowerment have different effects on the odds of domestic violence.

Results

Descriptive statistics for domestic violence, empowerment, community, demographic, and control variables are presented in Table 1. On average, 28% of women in Nepal reported ever experiencing domestic violence. In terms of decision-making in the household, women reported having a say in two to three decisions of the four options on average.

The multivariate results are presented in Table 2. This model provides the odds of ever experiencing domestic violence associated with household decision-making and attitudes towards wife beating. As predicted, women being empowered, measured in decision-making power within the household, correlates with significantly lower odds of experiencing IPV. For every household decision in which a woman has a say, the odds of ever experiencing IPV are reduced by 10%. This result is statistically significant.

The multivariate results are presented in Table 2. This model provides the odds of ever experiencing domestic violence associated with household decision-making and attitudes towards wife beating. As predicted, women being empowered, measured in decision-making power within the household, correlates with significantly lower odds of experiencing IPV. For every household decision in which a woman has a say, the odds of ever experiencing IPV are reduced by 10%. This result is statistically significant.

Being a member of the higher caste or living in urban areas are associated with increased odds of IPV when social economic factors are controlled for (Model 3). Region wise, the terai region shows significantly

Doll Number	Dress Color	Skin Color	Race or Ethnicity	Hair Color	Hairstyle	Eye Color
#1	Pink and White with Black Dots	Dark Brown	Black	Black with Brown Highlights	Ponytail	Light Brown
#2	Hot Pink	Tan	White	Strawberry Blonde	Ponytail	Light Brown
#3	Hot Pink	White	White	Black with Brown Highlights	Ponytail	Light Brown
#4	Yellow and White	White	Biracial; White and Black	Blonde	Ponytail	Green
#5	Pink and White with Black Dots	Medium Brown	Black	Light Brown	Ponytail	Light Brown
#6	Yellow and White	White	White	Blonde	Ponytail	Blue
#7	Yellow and White	Pale White	Asian (Japanese)	Black	Ponytail	Light Brown
#8	Blue and Pink	Tan	Latino	Black with Brown Highlights	Ponytail	Dark Brown
#9	Blue and Pink	Light Brown	Black	Black	Ponytail	Dark Brown
#10	Yellow and White	Tan	White	Dark Brown	Ponytail	Green
#11	Hot Pink	White	White	Honey Blonde	Ponytail	Light Brown
#12	Hot Pink	Brown	Black	Brown	Ponytail	Light Brown
	Pink and					

higher odds of domestic violence than the hill or mountain regions. The odds of Christians and Muslims ever experiencing domestic violence are about 80% higher than for other religions. The result is only significant when social economic factors are controlled for (Model 3). Among social and economic factors, the number of living children did not have a significant impact on the odds of ever experiencing domestic violence.

Empowerment is associated with reduced odds of domestic violence,

regardless of interactions between community variables and empowerment.

Discussion

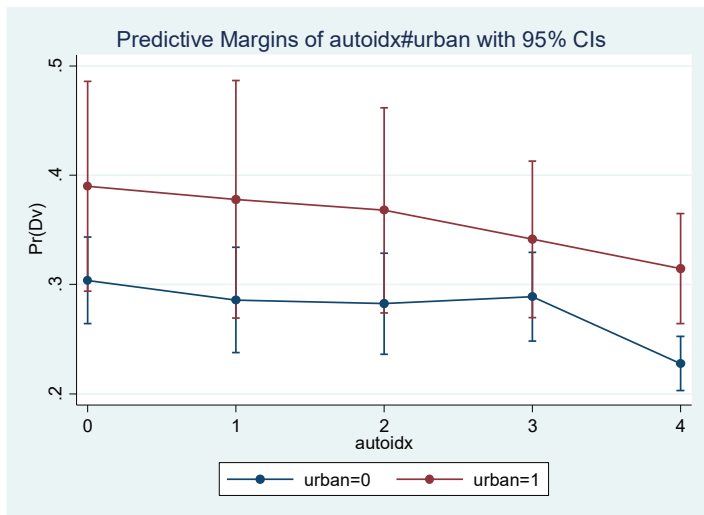
There are some important limitations to this study. Apart from the lack of longitudinal data to establish causality, the categories of this study are limited, and the sample size was not sufficient enough to capture the nuances of castes and religions in Nepal. For example, the survey did not distinguish between Tibetan Buddhism and Teradata

Buddhism. Geographic locators in this study did not go beyond ecological region level. Characteristics of the community, such as a village leader's attitude towards domestic violence, its governing structure, or the type of prevailing economic activity were not measured. Ethnographic research, if possible, should be conducted to examine the inclusion of these variables. The survey also did not include forms of domestic violence other than physical or sexual violence. Reports of domestic violence and

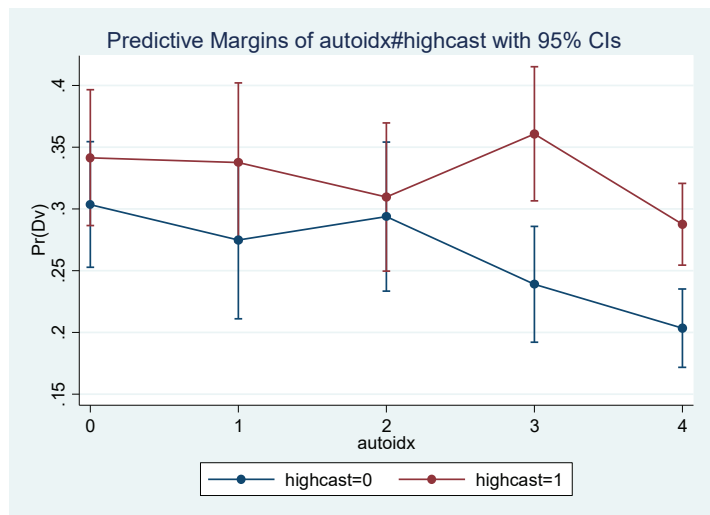
SOCIAL SCIENCES

Participant	Age	Race or Ethnicity	Doll Choice for Q1	Doll Choice for Q2	Doll Choice for Q3	Doll Choice for Q4	Doll Choice for Q5	Doll Choice for Q6	Doll Choice for Q7
#1	10	Biracial; Black and White	9	9	8	9	9	10	7
#2	6	White	6	3	2	11	3	9	10
#3	9	Black	12	12	9	12	12	11	1
#4	7	White	3	3	9	6	3	8	7
#5	10	Black	5	7	13	1	1	1	2
#6	8	Biracial; Black and White	5	12	1 and 9	5	5	4	1
#7	9	White	11	3	2	10	8	13	1
#8	9	Biracial; Black and Latino	5	8	9	5	5 and 8	3	1 and 7
#9	10	White	13	10	9	6	5	11	7
#10	4	Black	5	5	1	5	5 and 13	13	4
#11	4	Black	9	9	4	9	5 and 9	12	2
#12	8	Biracial; Black and Latino	1	1	6	5	5	5 and 8	7
#13	9	Biracial; Black and White	4	12	1	4	4	2	1
#14	10	Black	12	9	7	1	12	12	2
#15	6	Biracial; Black and White	9	9	1	4	4	6	2
#16	6	Black	12	5	11	9	9	3	11
#17	10	Latino	8	8	1	8	8	3	12
#18	10	Latino	8	8	1	5	5	9	2
#19	7	Asian (Indian)	5	5	6	1	1	6	11
#20	9	Asian (Indian)	5	5	1	5	5	7	2
		Biracial; Asian							

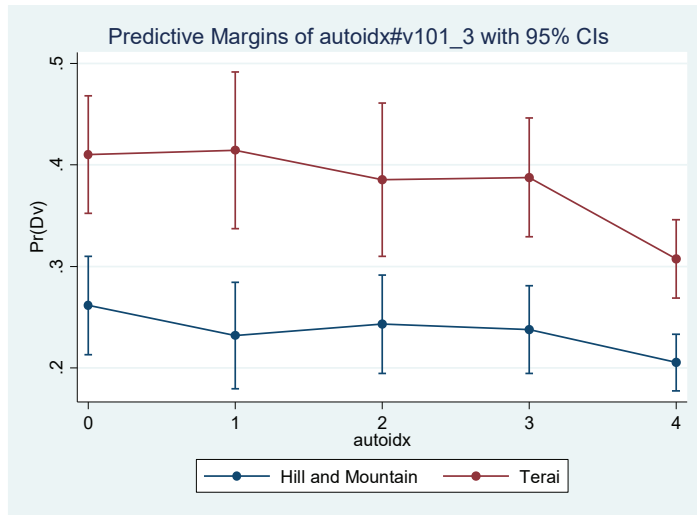
Graph 1



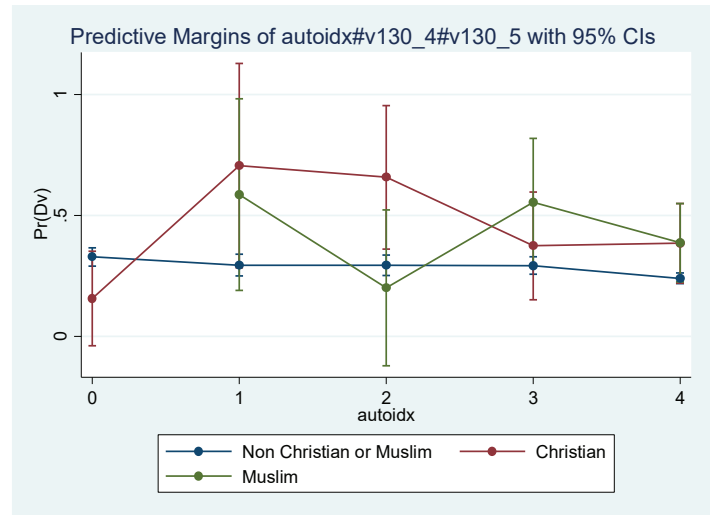
Graph 2



Graph 3



Graph 4



attitude towards wife beating had too few cases for a reasonable regression model. We also cannot rule out the differences in reporting domestic violence. Women of a higher caste or education level may be better able to identify, and be more willing to, report domestic violence. Further qualitative studies should be conducted in the field.

These limitations withstanding, this study made some important contributions to the literature. It was found that lower wealth, lack of education, younger age at first cohabitation, and having a job were associated with increased odds of domestic violence. These variables are evidence that an individual's characteristics, especially those that induce strain on the family, are correlated with higher odds of domestic violence. However, other stressors do not have a significant relationship with domestic violence. The number of living children does not significantly impact the odds of women experiencing domestic violence when other socio-economic factors are controlled. Increase in women's age is correlated with increased odds of domestic violence, but older women are less likely to be educated and have more stress-inducing responsibilities. These findings suggest that other theories might be needed to explain the odds of domestic violence.

Bourdieuian social theory might help explain some of these findings.¹³ In Bourdieuian tradition, the relatively powerless—in this case women—would challenge the powerful social roles which are embodied in the husband and other relatives. Any challenger to the established habitus must face pressure from all other actors, and those actors, generally relatives and

community members, are less likely to resort to violence as they employ different kinds of control, such as social exclusion and gossip.

Many of this study's significant findings concur with this perspective of domestic violence. Being a member of higher caste increased the odds of domestic violence. Being a member of the higher caste means more rigid social expectations upheld by traditional family roles. Members of higher caste might have held a less fluid conception of social norms and felt more threatened when the norms were challenged. Therefore, it is more likely for women in a higher caste to transgress against the norm, prompting violence. Weaker social ties also contribute to increased odds of domestic violence. As families sever social ties that kept behavior in check, women may lose buffers to domestic violence. Women in higher castes, free from the need to unite against stigmas, may form weaker social ties than women in lower castes. Weaker social ties might also explain why urban residences significantly increase the odds of domestic violence when wealth is being controlled. In an urban community with weaker social ties, men may lack other social means to control their wives' behavior, so they resort to violence.

Increases in domestic violence are seen as well among religious groups that have weaker social ties. Religion plays a large role in Nepalese social interactions, and religious minorities are often excluded from the social structures revolving around Hindu or Buddhist practices. Christian and Muslim religions are correlated with increased odds of domestic violence. These groups are often more recent immigrants from India and other South Asian countries with weaker communities

and fewer social connections in Nepal. Those who live in more isolated communities have higher odds of experiencing domestic violence, as do members of the dominant social group whose social order is being challenged.

This theory requires further research that produces qualitative insight into family and community structures. Ethnographic work on the dynamics of family life in South Asian or other developing countries can answer what a quantitative survey cannot.

There is research suggesting that empowerment might increase domestic violence in culturally conservative areas.⁹ However, the current study found that even in terai regions, where odds of domestic violence are higher than all other regions, higher level of female autonomy is associated with lowered odds of domestic violence. In urban areas where the odds of domestic violence are higher, autonomy is still significant. These results suggest that women's empowerment has a real impact on domestic violence. This can also be attributed to a shift in power within the family. When more decision-making responsibilities are delegated to their wives, husbands might feel less inclined to perform acts of domestic violence. There is also an observed change in women when they go through empowerment programs, as they learn to stand up for themselves.⁸ Previous studies theorized that on the family level, shifts in power induce conflict because husbands feel their power is being stripped away.⁴ Empowering women might cause husbands to feel insecure and consequently increase rates of domestic violence. However, it is only when the shift in power is nascent that husbands feel threatened; therefore, the negative effect of empowering

women is likely transitional. As new habitus are formed, domestic violence is no longer considered acceptable, and empowering women is the only way to changing the power dynamic in the field to form new habitus.

Out of the three groups of variables that influence domestic violence, women's empowerment was the one with the most significant policy implications. Individual and societal level changes lack clear policy solutions, but empowering women has been implemented across the globe for decades. Giving women more resources might lead to short term negative consequences, but giving women more say in household and social matters will lead to a change in the norm that benefits all. Policies and developmental projects should be designed to give women more say in political, economic, and social matters. Grass roots movements working towards empowering women might be especially effective, as one can expect less resistance to normative changes in gender roles. Strengthening communities around women in urban areas can foster a social norm against domestic violence. Empowering women has a pivotal role to play in the fight against domestic violence.

References

- ¹Ministry of Health and Population (MOHP) [Nepal], New ERA, and ICF International Inc. (2012). Nepal Demographic and Health Survey 2011. Kathmandu, Nepal: Ministry of Health and Population, New ERA, and ICF International, Calverton, Maryland.
- ²Sarkar, N. N. (2013). "The cause and consequence of domestic violence on pregnant women in India." *Journal of Obstetrics & Gynecology*. 33.3. Pg 250-253.
- ³Yount, K. M., & Carrera, J. S. (2006). "Domestic violence against married women in Cambodia." *Social Forces*. 85.1. Pg 355-387.
- ⁴Carlson, B. E. (1984). "Causes and maintenance of domestic violence: An ecological analysis." *Social Service Review*. 58.4. Pg 569-587.
- ⁵Malhotra, A., & Mather, M. (1997). "Do schooling and work empower women in developing countries? Gender and domestic decisions in Sri Lanka." *Sociological Forum*. 12.4. Pg 599-630.
- ⁶Wolf, M. (1984). "Marriage, family, and the state in contemporary china." *Pacific Affairs*. 57.2. Pg 213-236.
- ⁷Afridi, F. (2010). "Women's empowerment and the goal of parity between the sexes in schooling in India." *Population Studies*. 64.2. Pg 131-145.
- ⁸Naved, R. T. (1994). "Empowerment of women: Listening to the voices of women." *The Bangladesh Development Studies*, 22.2/3. Pg 155-178.
- ⁹Koenig, M. A., Ahmed, S., Hossain, M. B., & Khorshed Alam Mozumder, A. B. M. (2003). "Women's status and domestic violence in rural Bangladesh: Individual- and community-level effects." *Demography*. 40.2. Pg 269-288.
- ¹⁰Gellner, D. N. (2007). "Caste, ethnicity and inequality in Nepal." *Economic and Political Weekly*. 42.20., Pg 1823-1828.
- ¹¹Bates, L. M., Schuler, S. R., Islam, F., & Md. Khairul Islam. (2004). "Socioeconomic factors and processes associated with domestic violence in rural Bangladesh." *International Family Planning Perspectives*. 30.4. Pg 190-199.
- ¹²Visaria, L. (2008). Violence against women in India: Is empowerment a protective factor? *Economic and Political Weekly*. 43.48. Pg 60-66.
- ¹³Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste*. London: Routledge & Kegan Paul.