

The Proceedings of GREAT DAY 2011

SUNY Geneseo

Copyright © 2011 SUNY Geneseo
All rights reserved.
ISBN: 1475114001
ISBN-13: 978-1475114003

ACKNOWLEDGMENTS

The State University of New York at Geneseo

The Proceedings of the Fifth Annual GREAT Day

Geneseo Recognizing Excellence, Achievement & Talent is a college-wide symposium celebrating the creative and scholarly endeavors of our students. In addition to recognizing the achievements of our students, the purpose of GREAT Day is to help foster academic excellence, encourage professional development, and build connections within the community.

Publisher: Milne Library, SUNY Geneseo

Publishing Supervisor: Sonja Landes, Librarian, Milne Library

Editor: Stephanie Iasiello, Geneseo Class of 2012

Special Thanks to the Scholarly Communications Committee of Milne Library for undertaking this project.

Additional Thanks to:

Cyril Oberlander, Interim Director, Milne Library

Kerry McKeever, Associate Dean of the College

Patty Hamilton-Rodgers, GREAT Day Coordinator

GREAT Day logos and cover format have been used with the permission of their original designer, Dr. Stephen West.

GREAT Day is funded by the Office of the Provost

The GREAT Day Website: <http://great.geneseo.edu>

Table of Contents

ABSTRACTS	VI
ARIELLE ARONOFF	13
<i>REFLECTIONS ON THE MATRIARCHS ORCHARD</i>	
KARA JAMES.....	19
<i>MAKING AND FIRING HOPEWELL POTTERY</i>	
JAMES CLARK	24
<i>ARE YOUR CREDIT CARDS SAFE FROM ME?</i>	
TYLER MASSARO	36
<i>STABILITY ANALYSIS OF FITZHUGH – NAGUMO WITH SMOOTH PERIODIC FORCING</i>	
GREGORY ROLOFF & JESSICA BOSCH.....	51
<i>INVESTIGATION INTO THE COMBINED AFFECTS OF TUMOR NECROSIS FACTOR-ALPHA, TAMOXIFEN AND 5-AZACYTIDINE ON HTB-26 BREAST CANCER CELLS</i>	
MARCOS DAVILA	54
<i>FREE RADICAL - INTERACTIVE GAME DESIGN</i>	
KELSY COCOZZO	56
<i>TEACHING STUDENTS OF OTHER LANGUAGES</i>	
SIERRA HUNT.....	79
<i>A STORY OF ETHNIC GERMANS IN HUNGARY: FROM THEIR ORIGINS TO THEIR EXPULSION</i>	
KEVIN MULLER.....	98
<i>GOVERNMENT-TAXPAYER TENSIONS IN THE WAKE OF FIN 48 AND SCHEDULE UTP</i>	
NICK BECHT	125, 130
<i>TRANS-IDENTITY IN DJUNA BARNES' NIGHTWOOD</i>	
<i>CAT ON A HOT STREETCAR NAMED DESIRE: INTERROGATIONS OF FEMALENESS --Or, THE MAD HEROINE CODED AS HOMOSEXUAL--IN TWO PLAYS BY TENNESSEE WILLIAMS</i>	
SARAH GRAHAM.....	135, 138
<i>DEFINED BY THE HOME: HOUSING AND CLASS CONNECTIONS IN GEORGE ORWELL'S KEEP THE ASPIDISTRA FLYING</i>	
<i>A WORLD APART: SILKO'S SUPPORT OF A GYNOCENTRIC EXISTENCE IN GARDENS IN THE DUNES</i>	
DANIEL HART.....	141
<i>A NEW WAY FOR LIFE: STRIVING FOR POST-HUMANITY IN OCTAVIA BUTLER'S FICTION</i>	

CAITLIN KOWALEWSKI	148
<i>FAITH AND AUTHORITY IN EURIPIDES' MEDEA AND THE BIBLE</i>	
SALLY PROFETA	151
<i>THE SECOND SHEPHERDS' PLAY: "COLL'S SPEECH"</i>	
DANIEL SCHLANT	154
<i>THE RICHEST ARE NOT ALWAYS THE WEALTHIEST</i>	
JOSEPH WOLF	157
<i>STRUGGLING FOR AN ANSWER TO CAPITALISM: EVELYN WAUGH AND GEORGE ORWELL'S PESSIMISTIC APPROACH</i>	
REBECCA PULLANO	161
<i>KANG KEQING: FROM VICTOR TO VICTIM</i>	
STEVEN CARLSON	167
<i>CHINESE EXPANSION INTO THE WESTERN HEMISPHERE</i>	
JANNA CISTERINO	179
<i>ENGAGING THE DIASPORA: PROSPECTS AND PITFALLS</i>	
WILLIAM LABATE	198
<i>THE POLITICS OF ENVIRONMENTALISM IN MODERN DAY CHINA</i>	
JOHNSON LIN	217
<i>IRAN'S NUCLEAR PROGRAM</i>	
DEREK WENG	236
<i>RED, WHY, AND BLUE: A STUDY OF UPSTATE LOCAL POLITICS</i>	
ANDREW DEMOSTHENOUS	251
<i>PSYCHEDELIC RESEARCH RESURGENCE</i>	
JOEL INBODY	263
<i>SOMATIC EXPERIENCES AND THE SOURCE OF RELIGIOUS CONVICTION</i>	

Abstracts

Anthropology

Reflections on the Matriarchs Orchard

Arielle Aronoff

*Session Chair: Rose-Marie Chierici, Anthropology
Faculty Sponsor: James Aimers, Anthropology
Session 3-A*

Herbal medicine, native ecological design, and Judaism come together in a study that explores the links between women and natural healing. This presentation will focus on my experience as an intern at Kayam Farm, a Jewish Educational Farm in Reisterstown, Maryland. The history of the four Hebrew Matriarchs, Sarah, Rebekah, Rachel and Leah, is steeped with trials and tribulations, celebrations and tradition. My internship allowed for me to explore this history and develop an herbal guide to the Matriarchs Orchard. The orchard has been planted as a space for women to celebrate the cycles of life. Our implementation of the project seeks to maintain harmony between people, culture, and the Earth as well as an integration of mind body and spirit in order to maintain a healthy garden and healthy selves. This project has developed from the desire to connect contemporary people with the tradition of their ancestors in a symbolic gesture towards the agrarian lifestyle that was once the basis of the Hebrew culture.

Making and Firing Hopewell Pottery

Kara James

*Faculty Sponsor and Session Chair: Paul Pacheco, Anthropology
Session 2-A*

Between 200 BCE and 500 CE, Hopewell culture was flourishing across much of Eastern North America. They were living in dispersed sedentary households, and engaging in low-level food production, however, they are best known for building the large conglomerations of earthworks and mounds scattered across the Ohio Valley. Their impressive mounds have been excavated and documented, but not much is known about the domestic lives of these people. As archaeologists have started excavating Ohio Hopewell domestic sites, they have noticed that pottery not only is important in their burials, but also plays an important role in their daily lives as features associated with their dwellings contain many broken pieces of pottery. The pottery from both the domestic sites and the mounds has been classified, but archaeologists have not tried to recreate Ohio Hopewell pottery-making methods. This past semester, Geneseo student Kara James has been doing experimental research on the temperature firing ranges that the Ohio Hopewell used for their pottery production. This presentation is based on her research, findings, and methodology.

Applied Mathematics

Are Your Credit Cards Safe From Me?

James Clark

*Session Chair: Caroline Haddad, Mathematics
Faculty Sponsor: Gary Towsley, Mathematics*

I will begin with a brief overview of the history of cryptography and then specifically look at RSA cryptography. RSA is used for everything from secret communications, to wire transfers between banks, to transmitting your credit card information when you buy items online. I will go over three different factoring methods used to crack RSA cryptography - Pollard's Rho method, Fermat's method, and the continued fractions method. I will discuss the advantages and disadvantages of each, and conclude whether your credit cards are safe from me.

Stability Analysis of FitzHugh – Nagumo with Smooth Periodic Forcing

Tyler Massaro

*Session Chair: Caroline Haddad, Mathematics
Faculty Sponsor: Benjamin Esham, Mathematics
Session 3-M*

Alan Lloyd Hodgkin and Andrew Huxley received the 1963 Nobel Prize in Physiology for their work describing the propagation of action potentials in the squid giant axon. Major analysis of their system of differential equations was performed by Richard FitzHugh, and later on by Jin-ichi Nagumo who created a tunnel diode circuit based upon FitzHugh's work. The subsequent differential equation model, known as the FitzHugh-Nagumo (FH-N) oscillator, represents a simplification of the Hodgkin-Huxley (H-H) model, but still replicates the original neuronal dynamics. This thesis begins by providing a thorough grounding in the physiology behind the equations. We continue by proving some of the results postulated by Tanya Kostova et al. for FH-N without forcing. Finally, this sets up our own exploration into stimulating the system with smooth periodic forcing. Subsequent quantification of the chaotic phase portraits using a Lyapunov exponent are discussed, as well as the relevance of these results to electrocardiography.

Biology

Investigation into the Combined Affects of Tumor Necrosis Factor-Alpha, Tamoxifen and 5-Azacytidine on HTB-26 Breast Cancer Cells

Gregory Roloff and Jessica Bosch
Faculty Sponsor: Robert O'Donnell, Biology
Poster Presentation

The anticancer affects of Tumor Necrosis Factor Alpha and Tamoxifen have been widely recognized. Our preliminary experiments have revealed that HTB-26 breast cancer cells are nonresponsive to Tamoxifen due to low expression of the estrogen receptor. Other studies have shown that the 5 Azacytidine can stimulate the presence of the estrogen receptor. Experiments were carried out in an attempt to re-stimulate the sensitivity of breast cancer cells to Tamoxifen and measure the death-inducing properties of Tumor Necrosis Factor Alpha. Ideal results would include an increased susceptibility of Tamoxifen after a 5-Azacytidine treatment along with an increase in apoptosis, or programmed cell death.

Computer Science

Free Radical - Interactive Game Design

Marcos Davila
Faculty Sponsor and Session Chair: Christian Shin, Computer Science
Session 2-D

The program, titled Free Radical, is a video game that utilizes multiple List implementations to simulate a space shooter program in the vein of Space Invaders and Galaga. All enemies in this game are given a unique ArrayList that specifies their spots and the speed at which they move. Each enemy is given their own bullet, governed by the same list that "ghosts" the ship and becomes visible at a predetermined point, where it becomes visible and fires. Elements are shifted off-screen and "deleted" when the player's weapons intersect with an enemy. A specially scripted enemy known as the "overlord" is found at the end of each level, governed by its own list and its own methods and behaves differently compared to other enemies. Different MP3 music also plays in the background, which was made possible by clever manipulation of the bits of the music fed into the sound card using a byte array. The goals of this program were: to see if manipulation of lists and representation of manipulated lists by a pre-rendered image could provide enough illusion to simulate a retro arcade game, to see if rectangle collision was an adequate tool for hit detection, and to see if manipulation of array and byte values could send MP3 data files to a sound card.

Edgar Fellows Capstone

Teaching Students of Other Languages

Kelsy Cocozzo

Faculty Sponsor and Session Chair: Olympia Nicodemi, Edgar Fellows Honors Program and Mathematics
Session 1-E

According to the Department of Education (2010), 21% of school age students speak a language other than English. One in twenty of these students are not proficient in English. Though the number of English language learners (ELLs) is increasing, New York State does not provide any type of recommended model for teachers to follow in order to effectively teach these students. This project explores the most efficacious instructional methods for teaching English to students who are native speakers of other languages through a literature review and interviews with teachers of ELLs in the Rochester area.

Mentor: Annmarie Urso

A Story of Ethnic Germans in Hungary: From their Origins to their Expulsion

Sierra Hunt
Session 1-F

The origins of ethnic Germans in Hungary can be traced back to the 1700's, when the Ottomans were driven from Hungarian land and the Habsburg Empire promoted German migration to the unoccupied area. Known as the "Great Swabian Migration", these Germans set up various villages around the present day capital, Budapest. When the Treaty of Trianon reestablished European borders after World War I, the Germans suddenly found they were enclosed within the borders of a foreign country. This paper traces the history of ethnic Germans to Hungary and continues this history through the personal stories of two ethnic German families beginning in the 1930's. Unaware of the political events shaping their lives, these families dealt with devastation, death and imprisonment during World War Two, oppression during Soviet occupation, and finally expulsion from a land they had worked and lived on for over two hundred years.

Mentor: Thomas Greenfield

Government-Taxpayer Tensions in the Wake of FIN 48 and Schedule UTP

Kevin Muller

Faculty Sponsor and Session Chair: Olympia Nicodemi, Edgar Fellows Honors Program and Mathematics

Session 2-E

The U.S. federal income tax code is incredibly complex. Taxpayers do their best to make sense of ambiguous rules in order to figure out their fair share, but what happens when the law is unclear and the taxpayer has to decide whether or not to report a tax liability to the Internal Revenue Service? For corporations, the answer to this question is getting more and more complicated. Because of an accounting rule that took effect in 2006 and a new IRS policy that will affect large companies starting this year, many corporations will soon be required to tell the IRS about every dollar that could potentially be taxable based on different interpretations of the rules, even if the company's managers stand by their own interpretation and are willing to take the case to court. The IRS says the new policy will allow for more efficient auditing, businesses say the government is playing unfairly. Who is right, and how does all of this affect John Q. Taxpayer? This project takes a critical look at the implications of the policies from accounting and policy perspectives.

Mentor: Harry Howe

English

Trans-identity in Djuna Barnes' *Nightwood*

Nicholas Becht

Faculty Sponsor and Session Chair: Alice Rutkowski, English

Session 3-J: Transgender Theory and Literature- Part 2

In her novel, *Nightwood*, Djuna Barnes presents the reader with Dr. Matthew O'Connor, a character whose identity is so complex that he arguably has no fixed identity. On the surface, he is a cross-dressing, homosexual gynecologist with no medical license. Neil Miller's "Pioneers of Sexology," provides contextualizing information about the prevailing 'knowledge' and attitudes regarding non-normative gender identities and sexualities around the time that *Nightwood* was written. Of particular interest is the notion of sexual inversion, or of the female soul residing in the male body. In her work, Marjorie Garber examines "vestimentary transgressions" and the erotic desire(s) and discomfort which these transgressions cause, suggesting that the transvestite is "a crisis of "category" itself." Garber's ideas aid in an analysis of the ways in which Dr. O'Connor defies categorization, as well as the ways in which other characters respond to his various transgressions. Finally, Esther Newton's explanations of the sartorial

and drag systems in "Selection From Mother Camp" provide additional ways for analyzing Dr. O'Connor and his speech, which may qualify him as a figure of Camp. This application of gender and queer theory to Barnes' novel results in interpretations that are as fascinating as they are seemingly infinite.

Cat on a Hot Streetcar Named Desire: Interrogations of Femaleness --Or, the Mad Heroine Coded as Homosexual--in Two Plays by Tennessee Williams

Nicholas Becht

Faculty Sponsor: Graham Drake, English

Session Chair: Claire Littlefield

Session 4-C: Investigations of Gay and Lesbian Literature

Cat on a Hot Streetcar Named Desire: Interrogations of Femaleness --Or, the Mad Heroine Coded as Homosexual-- in Two Plays by Tennessee Williams Tennessee Williams utilizes heterosexual female protagonists with unstable identities in two of his major plays, "A Streetcar Named Desire" and "Cat on a Hot Tin Roof," in order to covertly represent and explore themes of male homosexuality and desire, they are Blanche DuBois and Maggie Pollitt, respectively. He does this because he is writing for a largely conservative and heterosexual audience of which anyone who might be homosexual would likely be in the closet, this is clearly evocative of Eve Kosofsky Sedgwick's notion of the "Open Secret". Through their involvement and presumably sexual partnerships with men of dubious heterosexuality, Blanche DuBois and Maggie Pollitt become aligned with homosexuals and associated with homosexuality. Williams employs elements such as Camp and subverted forms of the Patriarchal Gaze in order to further interrogate and trouble our sense of Blanche's and Maggie's femaleness.

Defined by the Home: Housing and Class Connections in George Orwell's *Keep the Aspidistra Flying*

Sarah Graham

Faculty Sponsor: William Harrison, English

Session Chair: Sarah Graham

Session 4-B: George Orwell's *Keep the Aspidistra Flying*: A British Novel of the 1930s

Focusing on the relationship between social class and living arrangements, George Orwell's *Keep the Aspidistra Flying* portrays textually 1930s class differences in London, England. That a character's class defines how he or she lives is made noticeably clear by Orwell, but his portrayal centers on the most central marker of class stratification: where each character lives. After World War I, there was a shortage of housing in London because little funding was available to build new houses. Slum clearing and updating was also put on the back burner and allowed to grow worse, thus allowing lower-class housing fall into greater ruin with

the arrival of the Great Depression. As Charles Mowat explains, “[a]ny estimate of the condition of Great Britain in the thirties is bound to be [...] affected by the continuing evidences of inequality between classes, including both those in work and the unemployed, in matters of income and capital, nutrition, the death rate, infant and maternal mortality, and shelter.” (emphasis added, 490). Using *Keep the Aspidistra Flying* as the primary text, I will attempt to show how Orwell argues that housing and class are dependent on each other, especially in regards to the exclusive nature of class, the size of personal space, ownership of personal space, levels of health, and freedom of sexuality.

A World Apart: Silko's Support of a Gynocentric Existence in *Gardens in the Dunes*

Sarah Graham

Faculty Sponsor and Session Chair: Gillian Paku,
English

Session 1-K: Identity I: Families

Leslie Marmon Silko is a Native American feminist and author of the 1998 novel *Gardens in the Dunes*, which focuses on the contradictions between repressed females in the patriarchal Western culture and self-determined Native American women. Her skillful storytelling expresses the power that can be drawn from being part of “collective sisterhood” that resists existing within the confines defined by patriarchy. Silko’s representation of repressive upbringings of the character Hattie under patriarchy, her positive portrayal of self-determined sexuality through the experiences of female Native American characters, and the assertion of the existence of a “collective sisterhood” outside the influence of Western culture that welcomes all females that embrace matriarchy lays the foundation for Silko’s stance of challenging accepted gender roles in a mainly patriarchal world.

A New Way For Life: Striving for Post-Humanity in Octavia Butler’s Fiction

Daniel Hart

Faculty Sponsor: Graham Drake, English
Session 4-C: Investigations of Gay and Lesbian Literature

Octavia Butler is one of the only writers of science fiction who identifies as a black lesbian. This genre of writing, usually dominated by the white male experience, finds new light as Butler’s characters display myriad unique perspectives and identities. The conflicts and issues her fiction delves into hold significance to the Queer community. This paper explores two of her novels, *Dawn* and *Adulthood Rites*, as well as her short story “Bloodchild,” elucidating upon connections Butler makes between her fiction and the dehumanization any given person may feel as a part of humanity’s need to maintain hierarchical structures of inequality. Though Butler writes from the social position of a black lesbian, her fiction suggests a displacement of social structure would be favorable to

any attempts to fix what humans have already established. Thus, Butler suggests the only solutions to human inadequacy are extinction or post humanization.

Faith and Authority in Euripides’ *Medea* and *The Bible*

Caitlin Kowalewski

Faculty Sponsor: Graham Drake, English

Session Chair: Caile Morris

Session 4-D: Ancient Authority: Western Humanities I

In this paper I argue that Medea, Jesus, and his Apostles all exist as either geographic or ideological foreigners in their respective cultures, and conflict with both social norms and human figures of authority. They instead place an emphasis on divine leadership, and their relationships with these divine figures have a direct impact on how they clash with figures of power on earth.

The Second Shepherds’ Play: “Coll’s Speech”

Sally Profeta

Faculty Sponsor: Graham Drake, English

Session Chair: Michael Mosher

Session 1-J: Beowulf, Boethius, and Bantering Shepherds: Voices from the Middle Ages

Throughout the sixteenth and seventeenth century, English theater was significantly influenced by the success of medieval mystery plays not only because they were enjoyed by all facets of society, but also because they contained socio-political commentary relevant to the period in which they thrived. More specifically, The Second Shepherds’ Play by the Wakefield Master was one of the most well-known medieval mystery plays of its time. A shepherd, Coll, makes a speech in the opening of the play that is particularly significant because it represents the dissatisfaction and woe that those in the peasant class suffered daily in a feudal hierarchical society. With closer examination, one can find specific literary devices used by the Wakefield Master in order to achieve greater appeal to his audience and succeed in portraying the hardship and oppression men like Coll endured. Undoubtedly, this medieval mystery play along with many others impacted greatly the future of professional English theater.

The Richest Are Not Always the Wealthiest

Daniel Schlant

Faculty Sponsor: Graham Drake, English

Session Chair: Michael Mosher

Session 1-J: Beowulf, Boethius, and Bantering Shepherds: Voices from the Middle Ages

In his 6th century work, *The Consolation of Philosophy*, the Christian philosopher Boethius offered

commentary on subjects such as fame, fortune, and happiness that seems radical in today's materialistic society. Boethius felt that money actually carries very little value, and is actually a burden upon those who amass vast quantities. Wealth provides nothing in the way of happiness, for true happiness is found within one's self and one's family. Boethius' thoughts would no doubt seem foreign to a vast majority of current society. The extremely wealthy are set upon a pedestal above the rest of society, and their status is envied and sought by the masses. While rare, the values Boethius wrote of are not nonexistent. The Larges, a couple in Canada who won the lottery, are proof of this. Content with their simple rural lives and the companionship they share, the couple donated ninety-eight percent of their \$10.9 million draw to local charities. They perceived happiness not in their new found wealth, but in each other. They saw money as a burden rather than a prize. Their act of generosity is Boethius' ideals personified in today's culture.

Struggling For An Answer to Capitalism: Evelyn Waugh and George Orwell's Pessimistic Approach

Joseph Wolf

Faculty Sponsor: William Harrison, English

Session Chair: Sarah Graham

Session 4-B: George Orwell's *Keep the Aspidistra Flying*:
A British Novel of the 1930s

The themes and styles of the Depression Era novels *Vile Bodies* (1930) by Evelyn Waugh and *Keep the Aspidistra Flying* (1936) by George Orwell appear to reflect an increasing distrust of capitalism as an economic and social structure for England and the world. These two authors come at the issue from differing sides of the British class divide. Waugh satirizes the frivolous actions, attitudes, and tendencies of wealthy British youths to show the impermanence of their class and situation in an unstable society. He also analyzes how these actions may or may not contribute to their downfall and ultimate erasure from the public consciousness. Orwell, on the other hand, criticizes capitalism from the beggared and dark battle Gordon Comstock fights. Gordon rejects opportunity and compliance with standard economic society in order to stand against what he calls the "Money God". This casts Gordon into deep despair that lasts until his ultimate reacceptance of the mainstream economic system. Although a distinct displeasure with capitalism is well vocalized in these novels, the issues raised are not resolved. Rather than articulating true alternative solutions or modifications that could be made, these authors only pessimistically embrace war as an inevitable final outcome.

History

Kang Keqing: From Victor to Victim

Rebecca Pullano

Faculty Sponsor and Session Chair: Tze-Ki Hon,
History

Session 2-J: *The Politics of Culture: The Cultural Revolution in China, 1966-76*

This paper presentation would focus on the part played by Kang Keqing in Communist China. Kang was a female commander in the Red Army during the Long March and leader in Mao's China who later became a victim of the Cultural Revolution.

Political Science and International Relations

Chinese Expansion into the Western Hemisphere

Steven Carlson

Faculty Sponsor Edward Drachman, Political
Science and International Relations

Session Chair: Jeremy Grace, Political Science and
International Relations

Session 1-O: *Political, Economic and Environmental
Dimensions of Trans-nationalism*

Chinese Expansion into Latin America In the recent history, and likely to continue well into the 21st Century, the Chinese have aggressively been working to secure energy supplies throughout Latin America. There are strategic and economic reasons for this action. However, it is not an ideological issue, China is willing to trade with whoever is willing to trade with them. I will analyze why this is occurring, how China is going about its acquisition, and what are the likely results of this expansion into the Western Hemisphere.

Engaging the Diaspora: Prospects and Pitfalls

Janna Cisterino

Faculty Sponsor: Jeremy Grace, Political Science
and International Relations

Session 1-O: *Political, Economic and Environmental
Dimensions of Trans-nationalism*

New developments in international migration, communication, and technology have facilitated the growth of transnational politics. The scale of international migration today is unprecedented, as is the ability of emigrants to maintain relationships with individuals in their home countries as well as with their home governments. Countries around the world have recognized the importance of these emigrants. Emigrant-sending states wish to tap into the resources of their diasporas, whether these resources be in the

form of financial contributions, knowledge and skills, or political support. Engaging the diaspora in order to extract resources is especially important for developing countries. There are a number of ways for countries to engage their diasporas, including, but not limited to, rhetorical appeals, consulting diaspora organizations, extending dual citizenship, and offering incentives for financial contributions. This project will assess the benefits and drawbacks of the different methods that states use to engage their diasporas.

The Politics of Environmentalism in Modern Day China

William Labate

Session Chair: Jeremy Grace, Political Science and International Relations

Faculty Sponsor: Tze-Ki Hon, History

Session 1-O: Political, Economic and Environmental Dimensions of Trans-nationalism

This paper examines the limits of environmental activism in People's Republic of China (PRC), an authoritarian one-party state. During the past few decades China has undergone rigorous economic development resulting in a series of negative environmental externalities. In response, Chinese citizens have begun to resist many government development projects that come at the expense of their local environment. In their struggle, Chinese civil society must navigate through several barriers that the PRC places on dissent in order to avert challenges to their regime. This paper examines successful and unsuccessful strategies that civil society uses to achieve the aim of environmental protection. It also examines how the PRC responds to these protests in what could become a potentially explosive issue. In order to examine this issue at the local regional and national levels this paper analyzes three case studies on significant environmental conflicts within China. Several policy prescriptions are issued in order to suggest ways that both the government and civil society can more easily facilitate environmental discourse.

Iran's Nuclear Program

Johnson Lin

Faculty Sponsor: Edward Drachman, Political Science and International Relations

Session Chair: Robert Goeckle, Political Science and International Relations

Session 4-I: Current Issues in International Security

Iran's nuclear program is a serious issue for the international community, particularly Israel. Iran has sought greater uranium enrichment, most likely to develop nuclear arms. Many in the international community believe that Iran's nuclear ambitions must be limited. Five possible options for slowing Iran's nuclear development are sanctions, pre-emptive strikes, regime change, clandestine operations (to sabotage nuclear facilities and assassinate Iranian

nuclear scientists) and to allow Iran to continue its nuclear development. These five options will be examined, and the pros and cons of the options will be analyzed.

Red, Why, and Blue: A Study of Upstate Local Politics

Derek Weng

Faculty Sponsor and Session Chair: Marilyn Klotz,

Political Science and International Relations

Session 2-N: American Public Policy Panel

The American political process is established upon a self-adjusting mechanism that weathers down opposing factions and streamlines myriad needs to form a cohesive solution that best addresses the public good. The State of New York is an exceptional place where geographic, economic social, and political diversity coexist in polarizing camps and distinctive patterns. This paper touches upon one of the most significant division in state politics today- the often contentious divide between the heavily Republican upstate and the solidly Democratic downstate. The purpose of this paper is to search for an casual explanation in how the lack of a competitive two party system in conjunction with a continual rural Republican dominance in upstate New York is ultimately harming its vitality and regional growth. I'm expecting to find discrepancies in the current Republican public policies in effect today and the unfulfilled constituency needs of the upstate region. Insights from a previous political internship at a local state senate campaign as well as a collection of recent data from various state agencies will be utilized in the paper.

Sociology

Psychedelic Research Resurgence

Andrew Demosthenous

Faculty Sponsor: Elaine Cleeton, Sociology

Session Chair: Steve Derne, Sociology

Session 1-T: Ritual, Religion and Spirituality: Studies in Transformative Experiences

By the 1960's, more than 1000 peer-reviewed research articles detailed the use of psychedelic compounds. Passage of the Controlled Substance Act of 1970 classified psychedelic substances such as LSD, MDMA, and Psilocybin as Schedule 1 drugs with no medical value, interrupting research into the application of psychedelic compounds for well-being. New studies from institutions such as New York University, Purdue University, Harvard University, and Johns Hopkins University are producing evidence that use of these drugs in combination with psychotherapy may have potential value for treating mental illnesses such as addiction, mood, affective, and psychotic disorders. This project is concerned with the effects of these drugs on well-being and the reasoning behind the scheduling of these drugs at the highest level. I

conducted six open-ended interviews with a snowball sample of college students about their recreational experiences with LSD, MDMA, and Psilocybin. Preliminary findings suggest that college students use these drugs as a tool for spiritual and emotional development. They report positive transformative experiences and improved relationships with family and friends. These findings support the continued expansion of research on psychedelic assisted psychotherapy for treatment of mental illness.

Somatic Experiences and the Source of Religious Conviction

Joel Inbody

Faculty Sponsor: Steve Derne, Sociology

Session Chair: Steve Derne, Sociology

Session 1-T: Ritual, Religion and Spirituality: Studies in Transformative Experiences

In an American Anthropologist article (*Metakinesis: How God Becomes Intimate in Contemporary Christianity*) (2010) Tanya Luhrmann highlights ways in which the body can play an integral part in creating conviction for modern day Evangelical Christians. In order to experience God in a personal, intimate way, believers may "identify... the presence of God (in) the body's responses" through "the absorbed state we call "trance"" and general "bodily and emotional experiences." From comments made during services at an Evangelical church, a study was conducted into whether or not somatic experiences (goose-bumps, tingles/prickling sensations, heat, etc.) could create conviction in the minds of believers. Interviews were conducted with fourteen attendees of Crossroads Christian Church in Elma, NY. It was found that ten out of fourteen individuals described one or more somatic experiences as a personal source of conviction for them. The importance of such things in creating a general sense of conviction varied. For three individuals, a somatic experience was their first evidence of the truth of Christianity. This research extends Lurhmann's conclusions in showing that even physiological reactions of the human body may play a role in creating a general sense of a higher reality in Evangelical circles.

Reflections on the Matriarchs' Orchard: An Alternative Understanding of Jewish Women and Environmentalism

Arielle Aronoff

Judaism is a religion based on agricultural values of caring for the earth in a symbiotic relationship. It is a theocentric religion, placing God as the source of all creation, which necessarily takes on the view that all of creation is sacred. From the beginning of the Genesis creation myth to modern day Jews, the text and its interpretations hold people to a standard above all other creations, yet they are inextricably bound to the earth for they are made of the same dust and out of the same will of the Holy Creator. Today there are people who take these teachings and apply them to a modern day lifestyle. Kayam Farm is a Jewish educational farm in Reisterstown Maryland devoted to the study and teaching of Judaism and Jewish values within an agricultural framework based on the sacredness of the natural world. There are many tools that Kayam uses to accomplish this task. My focus will be on one in particular: The Matriarchs' Orchard. The Matriarchs' Orchard is a project that seeks to maintain harmony between people, culture, and the Earth as well as an integration of mind body and spirit in order to maintain a healthy garden and healthy selves. This project has developed from the desire to connect contemporary people with the tradition of their ancestors in a symbolic gesture towards the agrarian lifestyle that was once the basis of the Hebrew culture.

In the Beginning... The Biblical Roots of Environmental Judaism

The Hebrew creation myth takes place in the Garden of Eden, which is thought to be located in the Fertile Crescent with in the Tigris- Euphrates river basin. This is where the Jewish religion began and spread throughout the Middle East. Those who practice Judaism now live in almost every country in the world. The story goes, "In the beginning God created the heaven and the earth... and God said let there be light: and there was light..." (Genesis 1:1). God went on to create the oceans and the land and all that dwell in and upon it, lastly God created man, Adam (pronounced Ah-Dahm in Hebrew). The word *Adam* shares the Hebrew root, *Adamah* meaning Earth as well as *Dam*, blood. This is no accident; many interpretations of the Genesis Creation myth focus on the inherent

interconnectedness humans have with the earth. In Genesis chapter two it is written, "the Lord God formed man (*Adam*) from the dust of the earth (*Adamah*)" (Benstein, 2006). God saw that Adam needed a companion and so he created Eve, in Hebrew *Chava*. The etymological meaning of *Chava* comes from the root "*chai*," meaning life. Thus Adam named her "because she was the mother of all living" (Hoffman, 2010). Judaism is based on a theocentric point of view; God created the world and placed humans on it with the command *le'ovda u'leshomra* (to till it and to tend to it) (Genesis 2:15). Here lies the basis of Jewish agricultural tradition: to till and to tend, also translated as to work and to guard or to serve and preserve. Humans are not owners of the world but Stewards of it. They are honored with the ability to utilize the earth's natural resources to plant and reap the harvests, yet they must also tend to the Earth as God's creation, as a sacred entity. Humans are *shomrei Adamah* (guardians of the Earth) who have the responsibility to maintain God's creation. "God said to Adam, 'See my works how good and praiseworthy they are? And all that I have created I made for you. Be mindful then that you do not spoil and destroy My world — for if you do spoil it, there is no one after you to repair it'"(Troster, 2007).

With the tasks set forth to work and to guard the earth come responsibilities written down as a set of laws in the Torah and other religious texts of Judaism. There are laws regarding the human relationship to one another, to the land to plants, to animals, as well as to God. *Tzedakah* (commonly translated as charity) comes from the root of *tzedek* (righteousness). It is a means of distributing wealth amongst all members of society. All people are obligated to give to those in need in an amount dependant on what one can afford (Isaacs, 2008). The Jewish people began as semisedentary shepherds living in tent villages across the Middle East. As the populations grew and warfare erupted, cities were established and a Kingdom was formed in modern day Israel. In this Kingdom laws were set in place to govern behavior. These laws were built up from a shared belief in God's commandments. The Torah is a composition of these laws as well as a written history of the Jewish

people. My focus will be on a particular set of laws regarding the distribution of agricultural yield to orphans, widows and those who do not own land.

Hebrew Law

One method of distribution to the poor is through *Pe'ah*, leaving the corners of your field for the poor to harvest. There are also two other practices of leaving food for the poor, *Leket* and *Shichecha*. This is the act of leaving any dropped produce or forgotten harvests, respectively, in the field for the poor to come gather. In addition, every third year the first fruits of the harvest, which would normally be paid as a tithe to the high priests, are commanded to go to the poor. This tithe of the first fruits is known as *Bikkurim*. On the years that were not designated as tithe to the poor, *Bikkurim* was brought to the holy temple in Jerusalem. It took place as a momentous gathering of farmers on their pilgrimage to Jerusalem where they were greeted with an elaborate ceremonial procession to the temple gates. The farmers entered into the temple and placed their offerings on an altar. This altar was the farthest point that anyone could go in the temple besides the high priests. The *Bikkurim* is thus interpreted by *Ishbitzer Rebbe*, *Mordechai Yosef Leiner* as symbolic of the holiness agricultural work possesses despite its classification as a devalued and disrespected position at that time (Vick, 2010).

There are as well laws to protect the land itself from human destruction. *Shmittah* (release), also known as *Shabbat Ha'aretz* (Sabbath of the Land), is the release of land from cultivation every seventh year. It is an important yet difficult law to keep, because the cultivation of land year after year leaves the soil barren of all nutrients and will continue to decrease in yield. However, a farmer must be sure to preserve an adequate amount of food from the previous year to live on in the sabbatical year. *Orlah* is a law that forbids the harvesting of fruits for the first three years from the time of planting the tree. The fourth year is a ritual celebration of offering the first harvests to God and not until the fifth year is the fruit available for the farmer to consume (Hoenig, 1969).

These instructions along with many other laws and regulations, address the notion of *Tikkun Olam* (Repairing the World), a concept based on the aforementioned notion of *Shomrei Adamah* (guardians of the Earth). We are the ones who must protect and tend to the earth because there is no one after us who can. It is our responsibility to leave the

world for future generations better than how we found it. This is the premise of Kayam Farm, a Jewish educational farm in Reisterstown Maryland. They are devoted to the study and teaching of environmentalism, farming and nature education within a Jewish framework.

When Traditional Meets Contemporary

The Torah is a book centuries old, and many practices are outdated. This leaves contemporary Jews with questions that cannot always be answered as we move through our daily lives. How is one to connect to ancient traditions when we can barely keep up with the world at hand? At Kayam Farm I felt grounded in a community devoted to merging the two worlds, ancient and contemporary. I am not the most knowledgeable person of Jewish theology and was never interested in exploring the ancient texts, however at Kayam I was. There I found a space where the texts were relevant. The laws became a foundation for a working organic, ethical, educational and spiritual farm where people are able to come for a day visit, a weekend retreat, a month long study and farming program, or for an entire seasonal internship.

Kayam is a community that is in a constant state of flux. There are full time workers who devote themselves entirely to programming, organizing, and maintaining the farm for others who come for a day, a week, a month, or a full growing season. Even those who come only for a day are just as much a part of the community as anyone. Each intern, volunteer, fellow, employee, and friend bring their own ideas and skills that aid in the evolution of Kayam. Jakir Manela, the founding director of Kayam holds a vision of an intentional Jewish farming community based on a shared sense of social and ecological responsibility.

Environmental Judaism uses Biblical texts as the basis for understanding the connection between humans and nature. The bible is not a set doctrine of beliefs, but a tool to be used to consider the complexities of the world in which we live. It can be seen as a multivocal figurative theoretical framework, which must be interpreted and put into practice. For example, laws such as *Bikkurim* and *Pe'ah* are brought into contemporary practice by donating a portion of the harvest to the Hannah More Family House, a local family shelter in the Baltimore community. In addition to social and ecological responsibility, Kayam is a place of religious practice. Distinctive spaces have been set as symbolic representations of Jewish traditions and

those devoted to learning and the growth of the individual. One such space is the Matriarchs' Orchard.

A Forest Garden

The Matriarchs' Orchard is a fledgling Permaculture Forest Garden, a sacred space for women to find a spiritual connection within themselves and to the earth through agricultural production as well as reconnecting with the past history and traditions of Jewish women. A Forest Garden is a food producing system that mimics the natural ecosystem of a forest. Most gardens as Americans commonly understand them are primarily composed of annual plants, each year an annual garden requires a large amount of manual labor including tilling, fertilizing, pruning, watering and so on throughout the season. A Forest Garden minimizes these inputs and increases the outputs of the plants by focusing on the production of primarily perennial and self-seeding plants. The basic premise of Forest Gardening is plant symbiosis, all plants work together in a beneficial community. Within the system a wide variety of trees, shrubs, vines and herbs exist, each one for a multitude of purposes both for human utilization and ecological health and stability.

A Woman's Sacred Space

The first task of this project is to clarify what determines a sacred space. Mircea Eliade speaks of the sacred and the profane as "two modes of being in the world" (Morris, 1987). He holds the position that every religious act and cult object is taken out of the world of the profane and placed into that of the sacred: "When a tree becomes a cult object, it is not as a tree that it is venerated, but as a *hierophany*, that is, a manifestation of the sacred... every religious act, by the simple fact that it is *religious*, is endowed with a meaning which... is 'symbolic' since it refers to supernatural values or beings" (Morris, 1987). In this context we must therefore refer to the symbolism of the Matriarchs' Orchard as an embodiment of Jewish values and the rich history of the Women who gave birth to the Hebrew people.

According to the *Kabbalah*, Jewish Mysticism, the *Shekhina* is the divine presence and imminence in the world. Many sects of Judaism interpret this as the feminine element of God who, in their interpretations, does not embody any specific gender. The *Shekhina* embodies the holy *sefirot*

(attributes of God), such as Wisdom and Understanding and Loving Kindness and, similar to the relationship between the moon and the sun, reflects these attributes to the earth. This representation of the *Shekhina* gives women a spiritual voice: it shows that they are not secondary characters that were composed as an afterthought from the rib of a man but complete spiritual beings made as well in the image of God (Prell, 2007).

The Matriarchs' Orchard is thus a symbol of the embodiment of the *Shekhina* in the natural world. It is a space for women to celebrate the cycles of life through traditional and modern rituals. The Mothers of Judaism, Sarah, Rebecca, Rachel and Leah are a source of inspiration and strength to contemporary Jews. In traditional Patriarchal Israelite society, men were the rulers of the household and women were expected to be subordinate to their desires however throughout the text women are described with valor. They are described with *sefirot* (attributes of God) that raise them to a holy stature:

According to the *Midrash* [Biblical Interpretation]... when Sarah lit *Shabbat* candles, they lasted through the entire following week; when she baked the *challah*, it too would remain fresh and warm for an entire week, from one *Shabbat* all the way to the next; and, for as long as she lived, a special cloud enveloped her and Abraham's tent. In this she provided spiritual and physical sustenance and protection to her home. When she dies, the candle goes out and the space she held in her home was changed – only when Rebecca filled that void was this space made complete and sacred again (Kayam Farm, 2011).

My role at Kayam was to establish meaningful connections between Sarah and the other Matriarchs to specific life cycle phases, such as birth, coming of age, pregnancy, and menopause. Through traditional Biblical translations, the stories of the Matriarchs are told and through various means of interpretation, myself and various other men and women at Kayam collaborated to assign each matriarch to a specific lifecycle. These were based on the highlighted events of their personal stories in the Torah. The dynamic nature of these events correlates with intense emotion and ritual celebration to mark the occurrence of each. The Matriarch's Orchard will provide a sacred space to mark such occasions as well as provide a source of nourishment for women during a time when the body is in need. Herbal medicine has been used for

centuries and is still the basis of many pharmaceuticals today. In this project I have researched complementary herbs to help women during times of physical stress.

Stories of the Matriarchs

Sarah, the first matriarch is honored as the mother of all Israelites, wife to Abraham and mother to Isaac. As described in the aforementioned *Midrash* she was a woman with a blessing of compassion and warmth. She welcomed all who appeared at her tent to rest and refresh themselves before continuing on their journeys. Despite her innate character to care for others, she was not able to conceive a child of her own. It was not until the age of 90 that an angel came to bring a blessing of a child. Sarah laughed at the thought of conceiving at such an old age (Genesis 18: 12-14). The angel's blessing came to fruition and Sarah gave birth to a son and named him Isaac.

By the age of 90 women have already gone through menopause and can no longer biologically conceive children. Sarah's mystical conception of Isaac will remain in mystery for it is under much scrutiny of Biblical scholars and is not the premise of this paper. It is said that Sarah lived to be the age of 127, and for this reason Sarah stands as the symbolic figure for the menopausal stage in a woman's life. Black Cohosh (*Cimicifuga racemosa L*) is a native to the eastern woodlands of America and Canada. The roots/rhizomes of the plant have been utilized to reduce the frequency and intensity of hot flashes associated with menopause and other symptoms connected to this change in a woman's reproductive cycle (Burdette et al., 2002).

The day of Sarah's death is marked as the day of Rebecca's birth. When Rebecca had come of age, Abraham's most trusted servant Eliezer was sent to find a wife for Isaac. On his journey, Eliezer devised a plan to find the right girl to bring to Isaac- one who showed compassion and selflessness. Immediately upon arriving at Rebecca's village he saw her bringing water from the well. Without hesitation, Rebecca offered the traveler water and brought plenty for his ten camels as well. Legend goes, when she went to fetch the water, it rose up to meet her as though she were the moon controlling the tides of the ocean. This is how Eliezer knew she was the right woman for Isaac. When Rebecca entered Sarah's tent she brought back the three miracles that were present while Sarah occupied it but which disappeared upon her death and thus brought joy and light once again in

to Abraham's village (Hammer, 2006).

Rebecca had to endure a long arduous journey from her home to that of Isaac's. As representative of this I have chosen the "walking onion" to symbolize her journey. *A. cepa proliferum* is a perennial which produces small bulbs on top of their stalks. When they mature, these stalks fall over planting new bulbs for the following year. This unique propagation causes the plant to "walk" around the garden. Rebecca embodies strength most notable among the four matriarchs. She endured a difficult pregnancy carrying twins, Esau and Jacob, who tradition tells have been fighting since their days in the womb. We therefore find it most helpful to relate Rebecca to pregnancy as a symbol of strength during this emotionally and physically charged period.

There are many herbs that help strengthen the body and work to relieve the pains associated with pregnancy. Red raspberry leaf contains fragrine, an alkaloid that gives strength to the muscles of the pelvic region including the uterus itself. This helps in reducing labor pains by toning and relaxing the uterus (McFarlin, 1999). Though nettles have a bad reputation for stinging and irritating the skin, the herb is quite beneficial. Nettle is a rich source of iron, calcium and vitamin A. It is used to enrich and increase the flow of mother's milk and to help relieve water retention. The sting from nettles only occurs in the plants' fresh state, cooking or drying the nettles will eliminate the stinging.

Rebecca's courage throughout her life journey is most marked in Genesis 27, when she uses trickery to deceive her blind husband Isaac into giving Jacob his blessing for the birthright of the eldest son even though it was actually Esau who came out of the womb first. This decision marks Rebecca's empowerment, though it may not be an outward rebellion against the patriarchy, in this way she establishes her position as a competent decision maker with the knowledge between right and wrong. To protect Jacob from Esau's rage Rebecca sent him away to her brother Laban's village to seek a wife from his daughters Leah and Rachel (Newsome, 1998).

Rachel, the younger of the two sisters, was set to marry Jacob, however Laban deceptively switched his daughters on the wedding day and married Leah to him first. The two sisters became co-wives and much rivalry was between them. Rachel in her youthfulness could not conceive, she was distraught at the site of her fecund sister Leah who gave birth to seven children. After much prayer and anticipation, Rachel gives birth to

Joseph and later dies while giving birth to Benjamin. The portrayal of Rachel as a young bride corresponds to our representation of her as illustrative of coming of age.

There are many herbal remedies to relieve the symptoms of menstruation. Yarrow (*Achillea millefolium*) is an herb used to reduce excessive menstruation and bleeding. It eases menstrual cramps and can help stimulate delayed or absent menstrual cycles. Cramp Bark (*Viburnum opulus*) is a uterine tonic and sedative with strong antispasmodic properties to relieve muscle aches and tension. Ginseng (*Panax quinquefolius*) is a plant noted for its ability to increase vitality and energy by acting on the pituitary and adrenal glands and stimulates the nervous system to decrease fatigue (Zak, 1999). These plants are symbolic of the youthful stamina and grace Rachel possessed and they are an aid to the current women of the Matriarchs' Orchard in order to maintain such stamina throughout their lives.

Leah, elder sister of Rachel and first wife of Jacob, symbolizes wisdom and wholeness. She is blessed with many sons and finds joy and gratitude in her good fortune despite her loveless wedlock to Jacob. Leah finds solace in her seven children six sons and one daughter. Her life giving presence is exemplified in her name. In *Gematria*, Jewish numerology, "Leah" adds up to 36. This number symbolizes double *chai*, or double life: *Chai*, life, being represented by the number 18 (Hammer, 2006). Leah's fecundity is symbolic of the modern day women at Kayam farm who are preparing to give birth. Partridgeberry (*Mitchella repens*) is used as an herb to prepare women for labor and soothe nervous irritability. It improves the tone and functional ability of a woman's uterus. Black Haw (*Viburnum prunifolium*) is another beneficial herb used during the latter stages of pregnancy to relieve false labor pains and after pains (Belew, 1999).

Conclusion

The Matriarchs' Orchard is currently in the beginning stages of development. The first tree was planted in 2007 and it has since grown into a productive orchard of peaches, apricots, persimmons, and many other fruit bearing trees and shrubs. It has notably been a place of ritual celebration for one woman in particular. Netsitsah Manela on two occasions held special tree planting ceremonies upon the birth of her two sons, Lev Yodea and Shama Nissim. Each tree was planted with the child's placenta underneath, which

contains abundant nutrients that add to the fertility of the soil and aid in the tree's growth. While Shama's apricot tree is less than two years old, Lev's peach tree has shown exceptional growth among the others. The peach tree was also the site of the ritual *upshirn* (a boy's first hair cut) when Lev turned three years old. Along with much dancing and celebration, Lev's hair was buried underneath his placenta tree and then the tree itself received its first pruning.

Today, many people are seeking to find a new sense of spirituality grounded in the natural ecology around them. We are seeing traditional text being transformed by new interpretations and given new meaning relevant to today's paradigm. Unique opportunities are emerging at places such as Kayam farm that acknowledge traditional Jewish values as a means to live in harmony with the environment. Women are placed in this context and celebrated for their unique spiritual connection to the *Shekhinah* and are accredited as powerful figures in a traditionally male-centric Biblical text. Humans have been set on the earth to till and to tend to God's creations, reaping the benefits of nature's bounty while ensuring the sacred nature of the earth is not squandered.

References

- Benstein, Jeremy
2006 The Way Into Judaism and the Environment.
Vermont: Jewish Lights Publishing
- Burdette Joanna E. , Chen Shao-nong , Lu Zhi-Zhen, Xu Haiyan, White Bethany E. P., Fabricant Daniel S., Liu Jianghua , Fong Harry H. S., Farnsworth Norman R., Constantinou Andreas I., van Breemen Richard B., Pezzuto John M., and Bolton Judy L.
2002 Black Cohosh (*Cimicifuga racemosa* L.)
Protects against Menadione-Induced DNA Damage through Scavenging of Reactive Oxygen Species.
Bioassay-Directed Isolation and Characterization of Active Principles: Journal of Agricultural and Food Chemistry 50(24):7022-7028.
- Belew, Cindy CNM, MS
1999 Herbs and the Childbearing Woman:
Guidelines for midwives. The American College of
Nurse-Midwives.
- Hammer, Jill
2006 The Jewish Book of Days: A companion for
all seasons. Philadelphia: The Jewish Publication
Society.

- Hart, Robert
 1996 Forest Gardening: cultivating an edible landscape. United Kingdom: Green Book Ltd.
- Hoffman, Joel M.
 2010 And God Said: how translations conceal the bible's original meaning. New York. St. Martin's press
- Hoenig, Sidney B.
 1969 Sabbatical Years and the Year of the Jubilee.
 The Jewish Quarterly Review 59(3): 222-236
- Isaacs, Ronald H.
 2008 Bubbe Meises: Jewish Myths Jewish Realities. Jersey City, New Jersey. KTAV Publishing House Inc.
- Prell, Riv-Ellen
 2007 Women Remaking Judaism USA: Cohn-Haddow Center for Judaic Studies
- Troster, Rabbi Lawrence.
 "Ten Jewish Teachings on Judaism and Environmentalism."
 2007. http://www.coejl.org/learn/je_tenje.php
 (accessed 3/23/2011).
- McFarlin, Barabra, Gibson, Mary, Harman, Patsy
 1999 A national survey of herbal preparation use by nurse-midwives for labor stimulation: Review of the literature and recommendations for practice:
 Journal of Nurse-Midwifery 44(3):205-216
- Morris, Brian
 1987 Anthropological Studies of Religion. New York: Cambridge University Press
- Newsom, Carol Ann & Ringe, Sharon
 1998 Women's Bible Commentary. Westminster: John Knox Press
- Vick, Yoni
 "Week Eight- Bikkurim". 9/1/2010.
<http://kayamfarmkollel.blogspot.com/2010/09/week-eight-bikkurim.html> (accessed 3/20/2011).
- Zak, Victoria
 1999 20,000 Secrets of Tea: The most effective ways to benefit from nature's healing herbs. New York: Dell Publishing

Making and Firing Hopewell Pottery

Kara James

Introduction



Between 200 BCE and 500 CE, a cultural horizon was flourishing across much of Northeastern North America. The people living at this time were living in dispersed sedentary households, and engaged in low-level food production [1]; however, they are best known for building the large conglomerations of earthworks and mounds that are all over the Ohio Valley. Their impressive mounds have been excavated and documented, but not much is known about the domestic lives of these people. As archaeologists have started excavating the domestic sites of the Hopewell, they have noticed that pottery is not only important in their burials, but also plays an important role in their daily lives. The features associated with their dwellings have many broken pieces of pottery within them. The pottery from both the domestic sites and the mounds has been classified, but archaeologists have not tried to recreate Hopewell pottery-making methods.

The objective of this research project was to determine the likely temperature range that the Hopewell used to fire their pottery. No kilns have been found associated with Hopewell domestic sites, so it is likely that they were using a type of outdoor fire. This project focused on two sets of variables with two options; a covered fire compared to an uncovered fire, and a pit fire compared to a ground fire. A comparison of colors between the experimental pottery and the Hopewell pottery will hopefully determine the likely temperature range that the Hopewell used to fire their pottery.

Ceramic Background

Before going on to my methodology, I would first like to address some issues inherent in this project that must be accepted, for the situation does not allow for us to work around them. These would be the challenges of re-creating ceramics, which include differences in the type of clay, and differences in the amount of inclusions.

Clay is formed by the gradual weathering of rocks; this means that the location where clay is formed leads to different types of clay [2]. If the clay has been formed in an area and has not been moved from that area prior to human collection, it is called primary clay. However, most clays are secondary clays – they were formed in one location via natural processes, and then transported from their location of formation to another location. Clay that is moved in this way picks up particles along the way, so the amount of carbonic acid and other diluted solvents is different in different clay deposits [2]. The clay that the Hopewell were using to create their pottery was likely found in the river, in a secondary clay deposit. This problem is compounded when we factor in time's affect on the situation; because rivers change their courses slowly over time, there is no way to get exactly the same clay as the Hopewell used. The different type of clay would affect the firing temperature, and it might affect the firing temperatures significantly. It is something we will have to accept as an assumption in this study.

Inclusions are mixed into clay to control the shrinkage of the clay – the change in size that occurs during water loss and firing [2]. They prevent the pottery from breaking in the creation process. However, the amount of inclusions in clay does not affect firing temperature; it only affects the functionality of the end product.

Methodology

All of the previous information was taken into consideration when I was selecting my materials; the focus of this study is a color comparison between the artifacts and experimental vessels, so it is incredibly important to obtain similar materials. In general, I tried to find as many materials as

possible from as similar a source as the Hopewell would have used. This way, the extent of the ceramic problems would be lessened.

The clay used in this project came from the banks of the Scioto River, which is the river that runs close to Brown's Bottom [1]. The exact location that the clay was acquired at was unknown. The inclusions used came from crushed local granite, which would have been plentiful near the Hopewell. In order to make the granite easier to crush, it was heated first.

Pottery Creation



The first step of vessel creation was to mix the clay and the inclusions together. This was done by breaking a larger chunk of clay into four smaller sections. Inclusions were then added to these smaller sections, and the smaller sections were recombined into the large chunk. This large chunk was continuously kneaded until the inclusions were evenly distributed throughout the clay. The distribution was observed by cutting the clay chunk in half occasionally throughout the kneading process. This combination of clay and inclusions was then used to make the coils for vessel creation.

Pottery vessels can be created many different ways; archaeological evidence supports the hypothesis that the Hopewell used a coil method to construct their vessels. The most substantial evidence



supporting this hypothesis is seen in the breakage patterns of the vessels that are excavated; oftentimes, they break along the neck in a fairly straight line, which would be right along the space between two coils. Based on this evidence, this experiment used the coil technique.

The combination of clay and inclusions, after being thoroughly mixed, is rolled into logs. These logs are stacked on top of each other in the desired shape of the vessel. The coils are then worked into each other on both sides to form the walls of the vessel. Next, tools were used to refine the vessel's shape, and increase its functionality.

Archaeological excavations of Hopewell sites have shown that they used some tools in their pottery production. They hit the sides of the pottery with paddles wrapped with lengths of hemp. This tool would help to better integrate the coils, which would produce a sturdier, more durable vessel. It was also likely used to regulate the surface of the pottery; the

Hopewell had large vessels, so this paddle would allow them to better balance the pot. On the insides of their pots, they used smooth stones to smooth the sides and bottom. They found these rocks in the nearby river. Only these two tools were used to reconstruct the vessels used in this study.

Firing

Before firing could begin, there had to be research on a few factors such as type of wood, method of measuring temperature, and the effect of environmental factors on fire temperature.

The fires in this study were fed primarily with hardwoods. Hardwoods are denser than softwoods, which makes them release more energy when they are burned. This makes them produce hotter fires than softwoods [3]. I assumed that since the Hopewell were making fires every day of their lives, they would have noticed that some woods burn hotter than others, and they would have used



these woods for firing their pottery. The wood used for the firing was mainly birch, with some other local hardwoods; these types of trees were found in forests near the Hopewell.

Research was also conducted on whether external environmental factors, specifically temperature, would have an effect on the temperature of the fire. No data or studies could be found on this subject, so it was assumed that the external temperature would not affect the fire's temperature. Moisture content in the wood would affect the fire's temperature and fuel efficiency, so all of the firewood was thoroughly dried beforehand.

The method of temperature recording used was a thermocouple probe. This device is used for measuring very high temperatures, and also would ensure fire safety. The probe is at the end of a two-foot long heat-resistant metal shaft; this shaft could be placed so that the sensor was in the center of the fire, while the handle at the other end would be safely out of reach of the fire. This sensor plugged into the thermocouple controller, and the controller would display the temperature of the area in both Fahrenheit and Celsius. Measurements were taken every half hour for the duration of the fire, and the first measurement was taken a half-hour after lighting the fire.

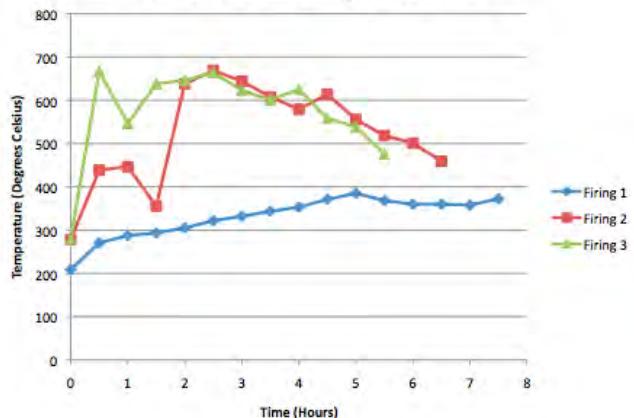
This experiment consisted of three different firings. The purpose of these three firings was to compare the changes in temperature in relation to two sets of variables: buried and unburied fires, and pit and ground fires. For the pit fires, the pit was 3-4 inches deep. The first firing was a buried pit fire, the second was an open pit fire, and the third was an open ground fire. These three firings allowed us to compare these two sets of variables and find out what variables have an effect on firing temperature. This would allow us to acquire a better understanding of the temperature range that the Hopewell used to fire their pottery.

Results

The data from all three fires was collected, and graphed on the same figure for the purposes of comparing the three fires. The graph is included below:

Some informal observations of the three fires

Temperature Change Over Time



should be noted here. The fuel ran out towards the end of the third firing, which caused the rapid decrease in temperature at the end of that firing. Temperature fluctuations were difficult to avoid for the second and third firings, because adding wood would cause the temperature to rise quickly, and then after a while the temperature would rapidly decrease. Also, the size of the wood placed on the fire seemed to have an effect on the temperature of the fire for the next half hour, causing more temperature fluctuations. These temperature fluctuations are unavoidable, and likely no attempt should be made to avoid them, because it is unlikely that the Hopewell would have evenly divided all their wood before adding them to the fire. The steadiness of the first firing was because it was buried; no additional fuel was being added, and the dirt acted as an insulation, therefore helping to maintain the temperature of the fire. It should also be noted that the temperature readings between the measurements for the third firing were significantly higher than the readings in the second firing; oftentimes, Firing 3's temperature would rise into the low seven hundreds. This would certainly affect the outcome of the ceramics from these two firings. The last notable point about the firing temperatures is that according to modern ceramics charts used by artists, none of these fires attained temperatures that would be hot enough to fully transform the pottery into ceramics [4].

After the experimental pottery was removed from the fire, it was brought to the lab so it could be broken and the colors could be compared to the Hopewell ceramics. The pottery from the first firing did not match any of the colors from the Hopewell pottery; in fact, there was little difference in color between the unfired pot and the fired pot. It was an orange/tan color. This pottery was also incredibly brittle. There are two possible explanations for why

this pot did not match any Hopewell pottery. The first explanation is that it is possible that the Hopewell were not using the pottery that was fired at a similar temperature range. Or, it is possible that the Hopewell did use pottery that was fired at this temperature range, but because the pottery is so brittle, none of the potsherds survived to be recognized during excavation. The vessels from the other two firings changed color significantly, and were also compared to Hopewell potsherds. The colors of the pots from the second and third firings did match some of the colors of the Hopewell ceramics. The pot from the second firing was a red/tan color, and the pot from the third firing was a red/dark gray color. Also, the third pot was the most difficult to break; however, all of these pots could be broken with less force than expected of ceramics. The pattern of color change seems to advance from low to high temperatures in this order: orange, tan, red, dark gray.



As the experimental pottery was being compared to the Hopewell pottery, it became apparent that the profiles of the experimental pottery changed color drastically. The color of the bases was a dark red, while the tops were a dark gray-black color; so the bases were exposed to a lower temperature range than the tops of all three of the vessels. It is possible that this color discrepancy

is only due to pottery placement in the fire, because all the vessels were placed top-up in the fire.

Discussion

These results give us some more information on the two variables' effect on firing temperature. For the comparison of the pit and ground fires, we can infer from the graph that the two pit firings did not get as hot as the ground firing. This phenomenon is likely occurring because the pit is reducing the amount of air circulating through the fire, even if it is a shallow pit. This reduction of air circulation would reduce the maximum temperature of the fire, and also reduce the fuel efficiency of the fire. The unburied pit fire consumed far more fuel than the ground fire, and left multiple large pieces of wood unconsumed. On the other hand, the ground fire reached significantly higher temperatures. This study suggests that it is likely that the Hopewell used a ground fire rather than a pit fire for ceramics firing.

We can also interpret the effect of burying a fire on its temperature. The buried fire did maintain a very steady temperature, but it did not come close to approaching the temperature of the unburied fires. This low temperature was produced by insufficient pottery; it did not change color much, and it was also very brittle. The two unburied fires achieved higher temperatures, but the temperatures did fluctuate a lot. There is evidence that significant temperature fluctuations weaken pottery, however, there is currently no archaeological evidence that would help us to figure out how much the Hopewell fires fluctuated as they made their pottery.

The last interesting subject is the brittleness of the experimental pottery. This may be connected to the size of the pottery found at the sites. During excavations, only incredibly small potsherds are found, and complete vessels are never found. It could be that the Hopewell pottery was very brittle, like the experimental pieces, and broke easily both before and after becoming a part of the archaeological record.

Conclusion

The findings from this study lead us to two possible conclusions. The first possible conclusion is that the color matching was not significant enough to conclude that the correct temperatures were achieved, and the Hopewell were likely using higher temperatures to fire their pottery. The

second possible conclusion is that the color matching was significant, which means the likely temperature range was determined, and that the Hopewell were not fully firing their ceramics. More research is needed to determine which of these conclusions will be better supported.

Further research is certainly needed, and this further research could take many different approaches. More tests could focus on the effects of burying a hotter fire, to see if that fire would maintain stable, high temperatures, and therefore fire the pottery more thoroughly. Or, possibly a bigger fire would achieve higher temperatures. It would also be wise to see if changing the orientation of the pot in the fire – such as placing the top down on the ground – would address the problem with the bases reaching lower temperature ranges than the tops of the pots. Also, performing studies of the vitrification of Hopewell pottery would help determine how much of the clay was converted to ceramics, which would give us a better idea of the likely temperature range used to fire their pottery.

References



[1]

Pacheco, Paul J, Burks, Jarrod, and Wymer, Dee Anne.

“The 2006 Archaeological Investigations at Brown’s Bottom 1.” *Current Research in Ohio Archaeology*
http://www.ohioarchaeology.org/joomla/index.php?option=com_content&task=view&id=26&Itemid=32

Pacheco, Paul J, Burks, Jarrod, and Wymer, Dee Anne.

“The 2007-2008 Archaeological Investigations at Lady’s Run.” *Current Research in Ohio Archaeology*
http://www.ohioarchaeology.org/joomla/index.php?option=com_content&task=view&id=28&Itemid=32

Pacheco, Paul J, Burks, Jarrod, and Wymer, Dee Anne.

“Investigating Ohio Hopewell Settlement Patterns in Central Ohio: A Preliminary Report of Archaeology at Brown’s Bottom 1.” *Current Research in Ohio Archaeology*
http://www.ohioarchaeology.org/joomla/index.php?option=com_content&task=view&id=103&Itemid=32

[2]

Foley, Nora.

1999 “Environmental Characteristics of Clays and Clay Mineral Deposits.”
<http://pubs.usgs.gov/info/clays/>

Blatt, H., Middleton, G., and Murray, R.

1980, Origin of sedimentary rocks (2d ed.): Englewood Cliffs, N.J., Prentice-Hall, 766 p.

Hillier, S.

1995 “Erosion, sedimentation and sedimentary origin of clays, in Velde, B., ed., Origin and mineralogy of clays: New York, Springer-Verlag, p. 162-219.

Velde, B.

1995 Composition and mineralogy of clay minerals, in Velde, B., ed., Origin and mineralogy of clays: New York, Springer-Verlag, p. 8-42.

[3]

The Wood Heat Organization Inc. Firewood.
2011. <http://woodheat.org/firewood.html>

[4]

Big Ceramic Store.com. Cone Temperature Chart.
<http://www.bigceramicstore.com/information/ConeChart.html>

[5]

Prufer, Olaf H.

1968 Ohio Hopewell Ceramics: An Analysis of Extant Collections. Anthropological Papers, Museum of Anthropology, University of Michigan No. 33. Ann Arbor, The University of Michigan

Are Your Credit Cards Safe From Me?

Cracking RSA Cryptography

James Clark

Introduction

We use cryptography every day. Governments use it to send secret communications, banks use it to send wire transfers, and we use it when we pay for anything with a credit card. If we want to send private information to anyone, cryptography is used in order to keep that information secret. So it is understandable that with such a high demand to send private information that is secure, it is important that the cryptosystem used cannot easily be cracked. In order to understand how we have reached this point, we shall look at the history of cryptography and how it has evolved to the systems we use today.

A Brief History and Evolution of Cryptography

The first cryptographic system widely used was the Caesar shift cipher. It was first documented as being used by Julius Caesar during the Gallic Wars to send military messages [10]. The idea behind this cipher is that the letters of the original message, or plaintext (represented by lower case letters), can be rewritten as different letters to create the encrypted message, or ciphertext (represented by upper case letters). An example of this is shown in Table 1. If we want to send the letter a, we would write D; to send b, we would write E; to send c, we write F; and so on and so forth. This example is the Caesar shift where the alphabet has been shifted by three places. Other examples of the Caesar shift can be seen in Table 2. The first one is an example where the ciphertext is completely random and the second one begins with the keyword or keyphrase Caesar Cipher (CAESRIPH) and then the rest of the ciphertext is alphabetical from the last letter of the keyphrase. Between these three types of Caesar shift algorithms, there are $26!$ permutations (or 403, 291, 461, 126, 605, 635, 584, 000, 000) of the Caesar cipher. Simon Singh notes in his book The Code Book that if we were able to check one key per second, it would take a billion times the lifetime of the universe to check every key in order to decipher the message [10]. Thus there became a need for cryptanalysts to break the cryptographer's

codes and the war between the cryptographers and cryptanalysts began.

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C

Table 1: Caesar cipher that is shifted three letters

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
G	Q	D	N	F	J	P	H	V	A	S	I	O	E	Z	B	C	X	L	W	R	Y	T	K	U	M

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
C	A	E	S	R	I	P	H	J	K	L	M	N	O	Q	T	U	V	W	X	Y	Z	B	D	F	G

Table 2: The first is an example of a random key while the second has the keyphrase *Caesar Cipher*.

The Caesar cipher was secure for centuries, until frequency analysis was created [10]. The idea behind frequency analysis is that certain letters are used more frequently than others and this fact can be used to crack the Caesar cipher. Let us assume that we are using the English language and have intercepted a message encrypted with the Caesar cipher. After examining a few texts in English, we are able to establish the frequency that each letter is used, such as in Figure 1. Since e is the most common letter (as seen by Figure 1), if the ciphertext's most common letter is Q then we can assume that $Q=e$. This pattern can be followed for the rest of the letters in the message until the key is found

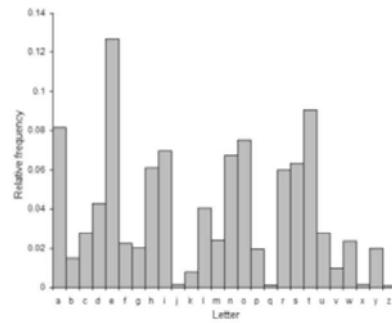


Figure 1: Frequency Analysis of the English Language

The next famous cipher was created in the early 1500s by a former diplomat Blaise de Vigenère[10]. His idea was to improve the Caesar cipher by encrypting the first letter with one Caesar cipher, the next letter with another Caesar cipher, the third letter with a third Caesar cipher, and so on and so forth. He created what is known as the Vigenère square (shown in Figure 2). A table and a keyword (such as *CODE*) would be agreed upon beforehand. To send a message, the first letter of the plaintext would be encrypted using the C row, the second letter with the O row, the third with the D row, and

the forth with the E row. Once the end of the keyword is reached the encryption would begin with the beginning of the code word again (in this case, with the letter C). The advantages of the Vigenère cipher is that with longer keywords it is more difficult to crack and it is also impervious to frequency analysis. Yet the cipher was eventually cracked in the 1850's by Edward Babbage because he was still able to find patterns [10].

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	
D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	
E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	
F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	
G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	
H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	
I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	
J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	
K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	
L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	
M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	
N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	
O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	
P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	
T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	
V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	
W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	
X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	
Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	

Figure 2: Virgenere Square

It was not until 1918 when the head of of cryptographic research for the U.S. Army, Major Mauborgne, created an uncrackable cipher using the Vigenère Cipher [10]. The idea behind the cipher is that a pad of random letters would be used as the keyphrase to encipher a message. Then once that message was sent the pad would be thrown away and a new pad would be used to encipher the next message. This cryptosystem, called the “one-time pad cipher”, has three advantages. The first one is that frequency analysis cannot be used since it is a type of Virgenere cipher. The second advantage is the number of keys that would have to be tested is extremely large. For instance, if one message had 21 letters for its keyphrase, then there are 26^{21} or 518, 131, 871, 275, 444, 637, 960, 845, 131, 776 permutations of the one-time pad key, which is more than any human or computer could test in a lifetime. The third advantage is best shown by an example. Let us say that we have intercepted this message from our enemies: PEFOGJJRNUKCEIYVVUCXL. Assuming that it is the length of the keyphrase is 21, we can either get the message attackthevalleyatdawn or defendthehillatsunsetor iwanttobeaagreendragon. Now we can assume the person sending the message is not talking about how he or she wants to be a green dragon, but we do not know if our enemy is going to

attack the valley at dawn or defend the hill at sunset. This shows us that we can get conflicting messages from the one-time pad cipher. In fact, we can get any message that is 21 letters long (assuming of course we are using a 21 letter keyphrase) which is what makes the one-time pad cipher uncrackable.

With the two world wars, there was a mechanization of cryptography with everything from decoder rings to the Enigma machine. All of these cryptosystems were eventually cracked except for the one-time pad cipher [10]. Yet the problem with the one-time pad cipher is that we have to send the key unencrypted, so we would either have to physically hand the key to the other person or send an unencrypted message that enemies could possibly intercept. Thus there was a need to be able to send encrypted messages without having to either communicate an encryption scheme with the other person ahead of time or send unencrypted keys that could potentially be intercepted.

Public Key Cryptography

As we saw with the one-time pad cipher, the problem is with the distribution of the key. If two people, let us say they are Alice and Bob, want to exchange messages, they must use a key which is a secret. So the problem becomes transmitting the secret key to the receiver in order to send the encrypted message safely. In other words, before the message is sent they must already have agreed upon a key.

Whitfield Diffie and Martin Hellman were working on the problem of public key cryptography and came up with a metaphor of what they wanted to accomplish. Let us say that Alice and Bob want to send a message to each other without Eve intercepting their message. Alice puts her message into a box, locks it with a padlock, and sends it to Bob. Bob cannot open it when he receives it because he does not have Alice's key, so he puts his own padlock on it and sends it back to Alice. Once Alice receives it, she unlocks her padlock leaving Bob's padlock to keep the box locked. She sends it back to Bob who can now unlock the box and read Alice's message. If Eve intercepts the box at any point that it is in transit between Alice and Bob, she will not be able to get into it [10].

There is a problem with this metaphor from a mathematical perspective that needs to be solved in order for the metaphor to work. The padlocks represents an encryption that is a one-way function,

which means that it is easy to compute forward and extremely difficult or impossible to go backward (or the inverse of the function is difficult/impossible to compute). So Alice first encrypts the message, then Bob encrypts the message, then Alice decrypts the message, then Bob decrypts the message so he can read it. The problem is that usually the last encryption needs to be the first one taken off, otherwise the message would not be readable [10]. The reason why the metaphor works though is because the padlocks are independent of each other. So they needed to find a one-way function that would work the same way the padlocks do on the box.

Diffie and Hellman solved the problem of key distribution in 1976 and outlined the solution in their paper “New Directions in Cryptography”. This Diffie-Hellman encryption used the fact that

$$Y = \alpha^X \pmod{q} \text{ for } 1 \leq X \leq q-1$$

is a one-way function for which it is extremely difficult to find its inverse. For instance, given any X we are able to calculate Y . Calculating X from Y where

$$X = \log_{\alpha} Y \pmod{q} \text{ for } 1 \leq Y \leq q-1$$

is much more difficult due to the difficulty of computing logarithms mod q , which is why the technique is so secure [4].

So let us assume that Alice and Bob want to send a message to each other. Each of them must generate a random number from the set $\{1, 2, \dots, q-1\}$ (X_A for Alice and X_B for Bob). Each of them are going to keep their respective numbers, X , a secret. Alice and Bob will communicate with each other and agree on an α and q . Next Alice will compute Equation 1 and Bob will compute Equation 2 and publish these results.

$$Y_A = \alpha^{X_A} \pmod{q} \quad (1)$$

$$Y_B = \alpha^{X_B} \pmod{q} \quad (2)$$

In order for Alice and Bob to communicate secretly, they calculate

$$K_{AB} = \alpha^{X_A X_B} \pmod{q}$$

and use that as their key.

Alice obtains K_{AB} by using Bob’s published number, Y_B and calculates

$$\begin{aligned} K_{AB} &= (Y_B)^{X_A} \pmod{q} \\ &= (\alpha^{X_B})^{X_A} \pmod{q} \\ &= \alpha^{X_B X_A} \pmod{q} = \alpha^{X_A X_B} \pmod{q}. \end{aligned}$$

Bob obtains K_{AB} in a similar way that Alice calculates K_{AB} ,

$$\begin{aligned} K_{AB} &= (Y_A)^{X_B} \pmod{q} \\ &= (\alpha^{X_A})^{X_B} \pmod{q} \\ &= \alpha^{X_A X_B} \pmod{q}. \end{aligned}$$

If Eve was able to intercept all of the information sent between Alice and Bob (α , q , Y_A , and Y_B), in order to know K_{AB} she would have to calculate

$$K_{AB} = (Y_A)^{(\log_{\alpha} Y_B)} \pmod{q}.$$

Diffie and Hellman point out that if logarithms mod q could be easily computed then the cryptosystem would not work. While they do not have a proof of the this fact (or its converse), they could not find a way to compute K_{AB} from Y_A and Y_B unless they either have X_A or X_B . Thus they have found a new public key distribution where only one key needs to be exchanged and its use can be coupled with a directory of user information to authenticate Alice to Bob and Bob to Alice [4]. The only disadvantage is that we still had to communicate the key to the person that we want to send the messages.

It was not until 1977 that Ronald Rivest, Adi Shamir, and Leonard Adleman created a public key cryptography in which the receiver and sender did not have to communicate outside of the encrypted messages. We can understand their process with another metaphor where Alice wants to send Bob a message. Again, she is going to put the message in a box, but instead of putting her padlock on the box she is going to put on a combination lock where Bob knows the combination. So she goes to the store and buys the “Bob Combination lock,” locks the box and sends the box to Bob. So when Bob receives the box, he can unlock it and read the message. Yet if Eve intercepts the box, she cannot get into it because she does not know the combination. So the lock itself would be the public key, the combination

would be the private key, the lock is the one-way function, and Alice and Bob never have to communicate outside the encryption. So how does RSA work?

Let us say that Alice wants to let other people, such as Bob, send her secure, encrypted messages that no one except Alice would be able to decrypt. She first selects two distinct primes, p and q , that are sufficiently large and random (e.g. not Mersenne primes) [3], where the two primes should not be close together, and both $p - 1$ and $q - 1$ have at least one large prime factor [6]. She then will multiply the primes together and call that number n (so $n = pq$). Alice then chooses a number e that is relatively prime to $\varphi(n)$ where $\varphi(n)$ is the Euler φ function and denotes the number of positive integers less than or equal to n that are also relatively prime to n (so $\varphi(n) = \varphi(p)\varphi(q) = (p-1)(q-1)$) [1]. $\varphi(n)$ and e are relatively prime if the greatest common divisor between $\varphi(n)$ and e is 1 (i.e. $\text{g.c.d.}(\varphi(n), e) = 1$) [1]. Once n is calculated and e is chosen, Alice has her public key and publishes these numbers in a directory. [3][6]

When Bob wants to send Alice a message, he wants to choose a ciphertext message that varies from user to user. So Bob wants to choose a plaintext to ciphertext algorithm that is uniform throughout the system. In order to do this he must first figure out what N -letter alphabet he wants to use (e.g. if he wants to use all of the letters, numbers, and symbols on the keyboard, then N would be 96.). He then chooses a $k \in \mathbf{N}$ such that $N^{k-1} < n < N^k$ (N^k should also be fairly large, Koblitz writes that it should be greater than 200 decimal digits [6]). Bob then takes the plaintext and splits it up into blocks of $k - 1$ letters, or $(k - 1)$ -digit base- N integers. This means that Bob assigns numerical equivalents between 0 and N^{k-1} for each $k - 1$ block. We similarly take ciphertext into units to be blocks of k letters in the N -alphabet. Therefore any plaintext message unit will be integers less than N^{k-1} and correspond to an element in $\mathbf{Z}/n\mathbf{Z}$ for any n . Since $n < N^k$, the image $f(P) \subset \mathbf{Z}/n\mathbf{Z}$ can be uniquely written as a k -letter block. It is important to note that only the corresponding k -letter block integers that occur are integers less than n , so not all of the k -block integers will be used. We transform the $k - 1$ block to a k block by taking each plaintext $k - 1$ block, P , and calculate $C = P^e \pmod{n}$ where C is the ciphertext block. This number C can then be written as a k block ciphertext. [6]

Alice can decode the message by computing the e -th roots of the digits received (and they are unique since e is relatively prime to $\varphi(n)$) [3]. Alice can do this by calculating d , where $ed \equiv 1 \pmod{\varphi(n)}$, by using Euclid's algorithm on e and $\varphi(n)$ since raising to the d -th power is the same as taking the e -th roots. In fact, Alice can compute d as soon as e has been chosen and does not have to remember p and q . Consequently, if Eve intercepts the message, she can decrypt the message only if she can factor n into its two primes since she can repeat Alice's process to calculate d . Therefore the message is secure if n cannot be easily factored.[3][6]

This method does allow Alice and Bob to send messages without communicating the key to each other, which Diffie and Hellman could not offer with their encryption scheme. Yet what is stopping Eve from sending a message to Bob pretending to be Alice? To solve this problem Alice needs to know Bob's public key (n_B , e_B) and her private key (n_A , d_A). If $n_A < n_B$, Alice would send Equation 3 in order to send her signature M . Bob can verify that the message is from Alice by calculating Equation 4 where C is the ciphertext, d_B is part of Bob's decipher key, and e_A is part of Alice's encipher key. If $n_A > n_B$, Alice would then send Equation 5 and Bob will compute Equation 6. [6]

$$C = [M^{d_A} \pmod{n_A}]^{e_B} \pmod{n_B} \quad (3)$$

$$M = [C^{d_B} \pmod{n_B}]^{e_A} \pmod{n_A} \quad (4)$$

$$C = [M^{e_B} \pmod{n_B}]^{d_A} \pmod{n_A} \quad (5)$$

$$M = [C^{e_A} \pmod{n_A}]^{d_B} \pmod{n_B} \quad (6)$$

Therefore the only way that Eve could read Alice's messages or even send messages as Alice is if she can factor n_A . So how hard can it be to factor n_A into its two prime factors?

Methods of Factorization to Crack RSA Cryptography

If we want to crack RSA cryptography we must be able to figure out the prime factorization of n . It is important to note that there are ways to manipulate the particular execution of RSA in certain cases that would allow Eve to learn the secret plaintext intended for Alice by using a chosen-ciphertext attack [7], but this is outside the scope of the project and we will be concentrating on methods of factorization to crack RSA cryptography. Thus, the most intuitive way to approach the problem is

to create a list of primes less than \sqrt{n} (e.g. Sieve of Eratosthenes or using primality testing on each number less than \sqrt{n}) and start dividing n by these primes to see if we get another prime q . Yet if n is sufficiently large it may take an extremely long time to arrive at n 's factorization. As a result, we will be looking at different factorization methods in order to find the prime factors of n .

Pollard's Rho Method of Factorization

The first method of factorization that we will be looking at is Pollard's rho method of factorization, which is the simplest factorization algorithm and substantially faster than the trial division of primes less than \sqrt{n} [6]. First, we choose an easily evaluated map $f : \mathbb{Z}/n\mathbb{Z} \rightarrow \mathbb{Z}/n\mathbb{Z}$, which means we choose a fairly simple polynomial mod n with integer coefficients. It is best that this function maps to itself in a disjointed manner (and should therefore not be linear and 1-to-1)[6], some examples of a function that we would use are $f(x) = x^2 + 1 \pmod{n}$ or $f(x) = 2x^2 + x + 5 \pmod{n}$. Next, we choose a particular starting values $x = x_0$, where x_0 can either be selected or randomly generated, and compute consecutive iterations of f (meaning that $x_{i+1} = f(x_i)$, for $i = 0, 1, 2, \dots$). Finally, we compare the different x_i 's in order to find two with different residue classes modulo n that are the same residue class modulo a divisor of n . When such an x_i and x_j are found, we compute the $\text{g.c.d.}(x_j - x_i, n)$ to find a prime factor n (meaning not 1 or n) and can then find the second prime factor by dividing n by the first prime factor. Since we are finding a point in the sequence where $x_i \equiv x_j$ we see that this pattern, which can be seen in Figure 3, resembles a Greek letter ρ and is where the method obtained its name. [6][8]

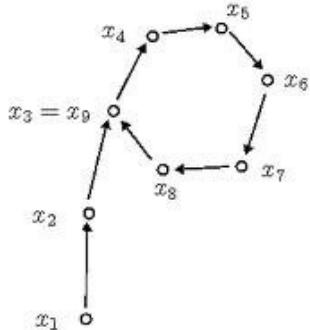


Figure 3: Where the Rho Method gets its name

Yet when j becomes substantially large, it is time consuming to compute $\text{g.c.d.}(x_j - x_i, n)$ for each $i < j$. We can minimize our work by computing one g.c.d. for each j . When we have found a i_0 and a j_0 such that $x_{i_0} \equiv x_{j_0} \pmod{r}$ for some r that is a divisor of n (i.e. $r|n$), we have the relation $x_i \equiv x_j \pmod{r}$ for any pair of indices i, j such that $j - i = j_0 - i_0$. So when $x_i \equiv x_j \pmod{r}$, then $x_{i+1} = f(x_i) \equiv f(x_j) = x_{j+1} \pmod{r}$ and the sequence x_i becomes periodic mod r with period $j - i$. One method then to reduce the number of g.c.d. 's that have to be calculated is after we compute x_j we suppose that j is an $(h + 1)$ -bit integer, which means that $2^h \leq j < 2^{h+1}$. Let i be the largest h -bit integer, which means $i = 2^h - 1$. We compute $\text{g.c.d.}(x_j - x_i, n)$ for these particular i, j . If the g.c.d. gives a nontrivial divisor of n then we have our answer, if not we make the same calculations for $j + 1$. Another method of calculating g.c.d. 's is to let b equal the period of the sequence (i.e. $b = j - i$), then $x_k \equiv x_l \pmod{r}$ whenever $k \equiv l \pmod{b}$, $k \geq i$, and $l \geq i$. Therefore, if we let k be the least multiple of b where $k \geq i$ and let $l = 2k$, then $x_k \equiv x_l \pmod{r}$. As a result, we can calculate $\text{g.c.d.}(x_{2i} - x_i, n)$ to find a nontrivial factor of n . [6][8]

The advantage of these two methods of calculating the g.c.d. is that we need to only compute one for each j , but the drawback to this is that it probably will not detect the first time there is a j_0 where $\text{g.c.d.}(x_{j_0} - x_{i_0}, n)$ is a nontrivial factor of n for some $i_0 < j_0$. Nevertheless, we will find a x_j and x_k whose difference has a common factor with n in less time. Also, the method may be expected to disclose the smallest prime factor p of n in roughly \sqrt{p} cycles, and is therefore faster than trial division for large n . It is important to note that the divisor obtained by this method may either be a prime factor, a composite number, or even n itself. In this situation when it is not a prime factor we would have to start over with a new x_0 , a new function f , a new way to compute the g.c.d. , or any combination of these three inputs. [6][8]

Fermat's Method of Factorization, Part I

The second method of factorization that we will look at relies on the theorem that every odd number in \mathbf{Z} can be represented by the difference of two squares ($\forall ac\mathbf{Z}, (a+1)^2 - a^2 = a^2 + 2a + 1 - a^2 = 2a + 1$). Thus we can represent the n from the public key in RSA cryptography as

$$n = pq = b^2 - c^2 = (b+c)(b-c).$$

Given this fact, we can let $p = (b+c)$ and $q = (b-c)$ [so $b = \frac{(p+q)}{2}$ and $c = \frac{(p-q)}{2}$].

$$b = \lfloor \sqrt{n} \rfloor + 1, \quad (7)$$

(where $\lfloor \cdot \rfloor$ is the floor function) and evaluate

$$b^2 - n = c^2. \quad (8)$$

If c^2 a perfect square, then we have found b and c and have therefore found p and q (since p and q defined in terms of b and c). If c^2 is not a perfect square, we evaluate $b = \lfloor \sqrt{n} \rfloor + 2$ and compute Equation 8 again to see if the new c^2 is a perfect square. We continue this iterated process until c^2 is a perfect square and therefore found our factors. [6]

If p and q are close together, then $c = \frac{p-q}{x}$ is small and b is slightly larger than \sqrt{n} . Therefore we will be able to find the factorization of the n after a few iterations of this process. So what happens if p and q are not close together? The Fermat method will eventually find p and q , but only after trying a large number of iterations of b , but there is a way to minimize the number of calculations and find our factorization. First, let $k \in \mathbf{N}$ be small and

$$b = \lfloor \sqrt{kn} \rfloor + 1. \quad (9)$$

We then evaluate

$$b^2 - kn = c^2 \quad (10)$$

and see if c^2 is a perfect square. We used the same process as before to find a b where Equation 10 is a perfect square. Once such a b is found, we know that $(b+c) = (b-c) = kn$ and therefore know that $b+c$ and n have a nontrivial common factor. We can then find this factor by computing $g.c.d.(b+c, n)$ to get one of our prime factors. This generalized Fermat method works quicker than the method outlined with Equations 7 and 8 when q is close to kp , and therefore reduces the number of b 's that need to be tried to find the factorization of n . This method can then be

generalized even further and leads to a more efficient factoring method. [6]

Fermat's Method of Factorization, Part II

An element $z \in \mathbf{Z}/n$ is a quadratic residue modulo n if there exists an $x \in \mathbf{Z}/p$ such that $x^2 \equiv z \pmod{n}$, and means that x is a square root of z . Also, if $n = pq$ were p and q are distinct primes, then every quadratic residue modulo n has exactly four square roots. Therefore, if we are given any $b^2 \equiv c^2 \pmod{n}$ where $b \not\equiv c \pmod{n}$, we can compute a non-trivial factor since Equation 11 implies Equation 12.

$$b^2 \equiv c^2 \pmod{n} \quad (11)$$

$$b^2 - c^2 \equiv 0 \pmod{n} \quad (12)$$

$(b+c)(b-c) \equiv 0$

We can then find a factor of n by computing $g.c.d.(b+c, n)$ or $g.c.d.(b-c, n)$ since $n|b^2 - c^2$ and $n \nmid (b+c)$ or $n \nmid (b-c)$ because $b \not\equiv \pm c \pmod{n}$. Therefore, $g.c.d.(b+c, n) = p$ is a proper factor of n and $q \frac{n}{p}$ is divisible by $g.c.d.(b-c, n)$. [5][6]

Thus, this quadratic sieve algorithm needs to find values for b and c such that their squares are equal modulo n . Yet if n is sufficiently large, a random selection of b where the least positive residue of $b^2 \pmod{n}$ is a perfect square is improbable and is necessary to generalize the method to allow flexibility in our choice of b 's to evaluate. Thus we want to choose several b 's with the condition $b_i^2 \pmod{n}$ is a product of small prime powers so that a subset of them will give a b whose square is congruent to a perfect square modulo n when multiplied together. [5][6]

So let B be a factor base, which is set $B = \{p_1, p_2, \dots, p_k\}$ of distinct primes (except p_1 may be -1). We then must find integers b such that, for any given n that we are factoring, $b^2 \pmod{n}$ is the least absolute residue and can be written as a product of numbers from B (found by trial division and primarily testing). We will call these numbers B -numbers. Also, b_i should be greater than \sqrt{n} to ensure that $b_i^2 \pmod{n} = \prod_{j=1}^h a_{ij}$ and the j th component of $\beta_i = \alpha_{ij} \pmod{2}$, so

$$\beta_j = \begin{cases} 0 & \text{if } \alpha_{ij} \text{ is even} \\ 1 & \text{if } \alpha_{ij} \text{ is odd} \end{cases}$$

So consider when we have some set of B-numbers b_i such that the corresponding vectors $\rightarrow_{B_i} = \{\beta_{i1}, \dots, \beta_{ih}\}$ add up to the zero vector in F_2^h . This means that the product of least absolute residues of b_i^2 equals the product of even powers of all of the p_j in B . So if for each i we let a_i be the least absolute residue of $b_i^2 \pmod{n}$ and write $a_i = \prod_{j=1}^h (p_j)^{\frac{1}{2} \sum_i \alpha_{ij}}$ we get

$$\prod a_i = \prod_{j=1}^h (p_j)^{\sum_i \alpha_{ij}} \quad (13)$$

$$= \prod_{j=1}^h \left[(p_j)^{\frac{1}{2} \sum_i \alpha_{ij}} \right]^2 \quad (14)$$

Since the exponent of each p_j in Equation 13 is an even number, we can rewrite the product as a square and get Equation 14. Hence, when we set

$$b = \prod_i b_i \pmod{n} \text{ and } c = \prod_j \left[(p_j)^{\frac{1}{2} \sum_i \alpha_{ij}} \right] \pmod{n},$$

where both b and c are least positive residues, we get two numbers whose squares are congruent modulo n . [5][6]

Of course, this fails if $b \equiv \pm c \pmod{n}$ and we must start over again with another collection of B-numbers whose corresponding vectors sum to zero (this will happen if we choose B-numbers less

than $\sqrt{\frac{n}{2}}$ since all of the vectors are zero-vectors and we get a trivial congruence). Yet if we choose b_i randomly we can expect that $b \equiv c \pmod{n}$ (up to ± 1) at most 50% of the time. The reason for this is because any square modulo n has 2^r square roots if n has r distinct prime factors. Therefore a random square root of b^2 has $\frac{2}{2^r}$ chance of being $\pm b$ (which is less than 50% since $r \geq 2$). So when we have $b^2 \equiv c^2 \pmod{n}$ where $b \not\equiv \pm c \pmod{n}$ we can find a nontrivial factor of n by computing $g.c.d.(b + c, n)$. Koblitz points out that if we go through this procedure of finding b and c until we find a pair that gives us a nontrivial factor of n , there is at most a 2^{-k} probability that it will take more than k tries. [6]

We can choose our factor base B and our B-numbers b_i with a few different methods. One method is to start with B consisting of the first $h - 1$ primes, let $p_1 = -1$, and choose random b_i 's until several are found that are B-numbers. Another method is to start choosing some b_i where $b_i^2 \pmod{n}$ is a least absolute residue and is small in absolute value (one way to do this is to choose b_i close to \sqrt{kn} where k is small). Then, choose B to consist of a small set of small primes (with $p^1 = -1$) so that several b_i 's are B-numbers when

$b_i^2 \pmod{n}$. It is important to note that given the collection of vectors in F_2^h , we can be sure to find a subset of them that sum to zero because we are looking for a collection of vectors that are linearly dependent over the field. We know from linear algebra that $(\rightarrow_{B_1}, \dots, \rightarrow_{B_h})$ is linearly dependent in F_2^h , if we have $h + 1$ vectors that are h -tuples. So we will have to generate at most $h + 1$ distinct B-numbers to find when

$$\left(\prod_i b_i \right)^2 \equiv \left(\prod_j (p_j)^{\frac{1}{2} \sum_i \alpha_{ij}} \right)^2 \pmod{n}. \quad (15)$$

If h is sufficiently large, it helps to write the vectors as rows in a matrix and use row-reduction techniques to find linearly dependent set of rows that sum to zero. [5][6]

We can now outline a systematic method of factoring n using Fermat's method and factor bases. First, choose an integer y of intermediate size (Koblitz gives the example of an intermediate size - if n is fifty digits, let y be five or six digits [6]). Let B be all of the primes less than or equal to y and -1 . Next, choose a large number of random b_i and try to express $b_i^2 \pmod{n}$ as least absolute residues that are products of the primes in B . Once we have obtained $(\pi(y) + 2) B$ -numbers (where $\pi(y)$ is the number of primes that do not exceed y [1]) we generate the corresponding vectors, (\rightarrow_{B_i}) , in F_2^h (where $h = \pi(y) + 1$) in order to get linear dependence. We then put the vectors into a matrix and use row-reduction to determine the subset of b_i 's whose corresponding \rightarrow_{B_i} add up to zero. We then let $b = \prod_i b_i \pmod{n}$ and $c = \prod_j (p_j)^{\frac{1}{2} \sum_i i \alpha_{ij}} \pmod{n}$ so that $b^2 \equiv c^2 \pmod{n}$ (which gives us Equation 15). If $b \not\equiv \pm c \pmod{n}$, compute $g.c.d.(b + c, n)$ to get a nontrivial factor of n . If $b \equiv \pm c \pmod{n}$, choose a different subset of rows in the matrix of \rightarrow_{B_i} 's that sum to zero, adding a few more B-numbers and their corresponding rows if necessary. If this does not work then we start the process over again with new random collection of B-numbers. [6]

Continued Fraction Method

The continued fraction method of factorization is a refinement of Fermat's method of factorization with factor bases. In the previous section we needed to find a reliable method of finding integers b where $1 \leq b \leq n$ such that the least absolute residue $b^2 \pmod{n}$ is a product of small primes,

which is likely to happen when $|b^2 \pmod n|$ is small. This method uses continued fractions, hence the name, to find b's such that $|b^2 \pmod n| < 2\sqrt{n}$. So we need to understand a few things about continued fractions before we can understand this factorization method.

A continued fraction is defined as follows: Given a real number x, we construct its continued fraction by first letting

$q_0 = \lfloor x \rfloor$ and set $x_0 = x - q_0$. Then let $q_k = \left\lfloor \frac{1}{x_{k-1}} \right\rfloor$ and $x_k = \frac{1}{x_{k-1}} - q_k$ until $\frac{1}{x_{k-1}}$ is an integer (and therefore $x_k = 0$). The process will terminate when x is rational because all x_k 's will be rational with decreasing denominators. Therefore, the continued fraction looks like

$$x = q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \dots + \cfrac{1}{q_k + x_k}}} \quad (16)$$

$$= q_0 + \cfrac{1}{q_1 + q_2 + \dots + \cfrac{1}{q_k + x_k}} \quad (17)$$

where Equation 17 is the compact notation of Equation 16 and each continued fraction is unique by the way we have defined it [3][6]. So the general form of a continued fraction is

$$q_0 + \cfrac{1}{q_1 + q_2 + \dots + \cfrac{1}{q_k}} \quad (18)$$

and we can let q_0, q_1, \dots, q_k be variables in order to manipulate the continued fraction into the expression of a quotient of two sums where each sum is composed of various products from the set $\{q_0, q_1, \dots, q_k\}$. We see that if $k = 1, 2, 3$ we can manipulate the continued fraction as

$$q_0 + \cfrac{1}{q_1} = \cfrac{q_0 q_1 + 1}{q_1} \quad (19)$$

$$q_0 + \cfrac{1}{q_1 + q_2} = q_0 + \cfrac{q_2}{q_1 q_2 + 1} = \cfrac{q_0 q_1 q_2 + q_0 + q_2}{q_1 q_2 + 1} \quad (20)$$

$$q_0 + \cfrac{1}{q_1 + q_2 + q_3} = q_0 + \cfrac{q_2 q_3 + 1}{q_1 q_2 q_3 + q_1 + q_3} = \cfrac{q_0 q_1 q_2 q_3 + q_0 q_1 + q_0 q_3 + q_2 q_3 + 1}{q_1 q_2 q_3 + q_1 + q_3} \quad (21)$$

where the value of $q_1 + \frac{1}{q_2}$ in Equation 20 is derived from Equation 19 and evaluating q_1 and q_2 in place of q_0 and q_1 respectively. Equation 21 was derived in the same way using Equation 20. Thus, we can build out the general continued fraction in this manner and can represent the numerator of Equation 18 as $[q_0, q_1, \dots, q_k]$. When we look at Equations 19, 20, and 21 we see that the denominator of the expression can be represented as $[q_1, q_2, \dots, q_k]$ since the denominator is derived from the numerator of the previous answer. Therefore, we can represent the general continued fraction as

$$q_0 + \cfrac{1}{q_1 + q_2 + \dots + \cfrac{1}{q_k}} = \cfrac{[q_0, q_1, \dots, q_k]}{[q_1, q_2, \dots, q_k]}.$$

We can also see from Equations 19-21 how the numerator is derived as the recurrence relation

$$[q_0, q_1, \dots, q_k] = q_0[q_1, q_2, \dots, q_k] + [q_2, q_3, \dots, q_k] \quad (22)$$

This will work for all k provided that when $k = 1$ we interpret the second bracket as 1 ($[q_0, q_1] = q_0[q_1] + 1 = q_0 q_1 + 1$, which is what it is supposed to be). [3]

So we see that $[q_0, q_1, \dots, q_k]$ is the sum of certain products formed from the set $\{q_0, q_1, \dots, q_k\}$, but how do we form the products? The answer was found by Euler, who was the first to give a general account of continued fractions, and created "Euler's rule" to form the products. The rule instructs us to first take the product of all the terms, then take every product that can be acquired by omitting any pair of consecutive terms, then take every product that can be acquired by omitting any two separate pairs of consecutive terms, and so on and so forth. The summation of these products is equal to the value of $[q_0, q_1, \dots, q_k]$. If n is odd (giving us $n + 1$ terms, which is even) we add the empty product or the product of all the terms omitted, which we have defined as 1. This can be proved by induction using the recurrence relation in Equation 22. We assume the rule holds for the right hand side of Equation 22 and we have to prove that it holds for the left hand side. $[q_2, q_3, \dots, q_k]$ is the sum of all the products formed from the set $\{q_0, q_1, \dots, q_n\}$ where q_0 and q_1 have been omitted. Therefore $q_0[q_1, q_2, \dots, q_k]$ is the sum of all the products formed from the set $\{q_0, q_1, \dots, q_n\}$ where q_0 and q_1 is not omitted since all the products must contain q_0 . When that factor is removed we are left with the sum of products from the set $\{q_1, \dots, q_n\}$ where any separate pairs of consecutive terms are omitted. Thus, we get the appropriate sum of products and the rule holds for the function $[q_0, q_1, \dots, q_k]$ for all k. [3]

Davenport points out that an immediate deduction of Euler's rule is that the value of $[q_0, q_1, \dots, q_k]$ is the same if the terms are rewritten the reverse order, meaning $[q_0, q_1, \dots, q_k] = [q_k, q_{k-1}, \dots, q_0]$. This is true because we can express $[q_0, q_1, \dots, q_k]$ in terms of a similar function as Equation 22, except with the last term or last two terms omitted:

$$[q_0, q_1, \dots, q_k] = q_n[q_0, q_1, \dots, q_{k-1}] + [q_0, q_1, \dots, q_{k-2}] \quad (23)$$

This is equivalent to Equation 22 because we can rewrite the terms in the opposite order to get $[q_k, q_{k-1}, \dots, q_0] = q_k[q_{k-1}, \dots, q_0] + [q_{k-2}, \dots, q_0]$ which is a restatement of Equation 22 with different symbols. [3] We can now define the i -th convergent of the continued fraction as the continued fraction terminated at q_i where $i < k$ of

the general continued fraction (Equation 18). The value of the i -th convergent is

$$q_0 + \frac{1}{q_1 + \frac{1}{q_2 + \dots + \frac{1}{q_i}}} = \frac{[q_0, q_1, \dots, q_i]}{[q_1, q_2, \dots, q_i]} = \frac{A_i}{B_i},$$

where $A_i = [q_0, q_1, \dots, q_i]$ and $B_i = [q_1, q_2, \dots, q_i]$ to simplify notation. We see from this definition that the first convergent is $\frac{A_0}{B_0} = \frac{q_0}{1}$, the last convergent $\frac{A_k}{B_k}$ is the continued fraction itself, and A_i and B_i for $0 \leq i \leq k$ are natural numbers defined by the sums and products of q_i 's as we have defined above. The recurrence relation from Equation 23 can now be represented as

$$A_i = q_i A_{i-1} + A_{i-2} \quad (24)$$

$$B_i = q_i B_{i-1} + B_{i-2} \quad (25)$$

where B_i has q_0 omitted. [3] From these two equations, we can show that any two consecutive convergents will differ by a factor of $(-1)^{i-1}$, or

$$A_i B_{i-1} - B_i A_{i-1} = (-1)^{i-1}, \quad (26)$$

which we can prove by induction. When $i = 1$, we have: $A_0 = q_0$, $B_0 = 1$, $A_1 = q_0 q_1 + 1$, and $B_1 = q_1$. So $A_1 B_0 - B_1 A_0 = (q_0 q_1 + 1)(1) - (q_0)(q_1) = 1$. We can prove this generally by substituting Equations 24 and 25 into Equation 26 to get

$$\begin{aligned} A_i B_{i-1} - B_i A_{i-1} &= (q_i A_{i-1} + A_{i-2}) B_{i-1} - (q_i B_{i-1} + B_{i-2}) A_{i-1} \\ &= A_{i-2} B_{i-1} - B_{i-2} A_{i-1} \\ &= -(A_{i-1} B_{i-2} - B_{i-1} A_{i-2}). \end{aligned} \quad (27)$$

Therefore, the expression on the left of Equation 26, which we will call δ_i , has the property that $\delta_i = -\delta_{i-1}$ which can then be recursively defined to give us $\delta_i = -\delta_{i-1} = +\delta_{i-2} = \dots = \pm \delta_i$ where $\pm \delta_i$ is $+1$ if i is odd and -1 if i is even. Therefore, it can be represented by $(-1)^{i-1}$ due to the fact that $\delta_1 = 1$ and Equation 27 is true. [3] A consequence of Equation 26 is that A_i and B_i will always be relatively prime and therefore the fraction B_i is in lowest terms, which means that all convergents (including the continued fraction itself) is in lowest terms. So when we represent a rational number $\frac{a}{b}$ as a continued fraction, the convergents of the continued fraction compose a sequence of rational numbers with the last one being $\frac{a}{b}$. We can see that the convergents will alternately be less than and greater than the final value of $\frac{a}{b}$ by rewriting Equation 26 as

$$\frac{A_i}{B_i} - \frac{A_{i-1}}{B_{i-1}} = \frac{(-1)^{i-1}}{B_i B_{i-1}}. \quad (28)$$

Furthermore, since $B_i < B_j$ for $0 \leq i < j \leq k$ the difference in Equation 28 will decrease as i increases. We can also see that the even convergents will be less than $\frac{a}{b}$ the odd convergents will be greater than $\frac{a}{b}$, and that each convergent

will be closer to $\frac{a}{b}$ than the previous convergent.

[3]

In order to see the continued fraction method of factoring, we need to use these facts to prove that $|b^2 \pmod n| < 2\sqrt{n}$. First, if $x > 1$ is a real number whose continued fraction expansion has convergents $\frac{b_i}{c_i}$, then $|b_i^2 - x^2 c_i^2| < 2x$. Since x is between $\frac{b_i}{c_i}$ and $\frac{b_{i+1}}{c_{i+1}}$ and the absolute value of the difference between the successive convergents is $\frac{1}{c_i c_{i+1}}$ (from Equation 28) we can see that

$$\begin{aligned} |b_i^2 - x^2 c_i^2| &= c_i^2 \left| x - \frac{b_i}{c_i} \right| \left| x + \frac{b_i}{c_i} \right| \\ &< \frac{c_i^2}{c_i c_{i+1}} \left(x + \left(x + \frac{1}{c_i c_{i+1}} \right) \right). \end{aligned}$$

Thus,

$$\begin{aligned} |b_i^2 - x^2 c_i^2| - 2x &< 2x \left(-1 + \frac{c_i}{c_{i+1}} + \frac{1}{2x c_{i+1}^2} \right) \\ &< 2x \left(-1 + \frac{c_i}{c_{i+1}} + \frac{1}{c_{i+1}} \right) \\ &< 2x \left(-1 + \frac{c_{i+1}}{c_{i+1}} \right) = 0. \end{aligned}$$

So if n is a positive integer which is not a perfect square, we can let $\frac{b_i}{c_i}$ be the convergents in the continued fraction expansion of \sqrt{n} so that the residue of $b_i^2 \pmod n$ is the smallest in absolute value (between $-\frac{n}{2}$ and $\frac{n}{2}$) is less than \sqrt{n} . When we apply the previous theorem with $x = \sqrt{n}$ we see that $b_i^2 \equiv b_i^2 - nc_i^2 \pmod n$ and the latter integer is less than \sqrt{n} in absolute value. This is the key to the continued fraction method of factorization because we can find a sequence of b_i 's that have small residues when squared by taking the numerators of the convergents of the continued fraction expansion of \sqrt{n} . We do not even have to find the actual convergent since we are only using the numerator $b_i \pmod n$ and the fact that the numerator and denominator of the convergents become large is not a problem since we are working with integers less than or equal to n^2 when we are multiplying integers modulo n . [6]

So here is the algorithm for the continued fraction method of factoring. Let n be the integer that we want to find its factorization. We first set $b_{-1} = 1$, $b_0 = a_0 = \lfloor \sqrt{n} \rfloor$ and $x_0 = \sqrt{n} - a_0$ and compute $b_0^2 \pmod n$ [or $b_0^2 - n$]. Next, for $i = 1, 2, \dots$ we first set $a_i = \lfloor \frac{1}{x_{i-1}} \rfloor \pmod n$, then set $b_i = a_i b_{i-1} + b_{i-2} \pmod n$ and compute $b_i^2 \pmod n$. After this is done for several i , we look at the $b_i \pmod n$ which factor into \pm a product of small primes. We then create a factor base B that are

composed of -1 and the primes that occur more than once of the set $b_i^2 \pmod{n}$ or occur with an even power in just one $b_i \equiv 2i \pmod{n}$. We then list all of the numbers $b_i^2 \pmod{n}$ which are B -numbers, along with their corresponding β_i of zeros and ones and find a subset whose vectors sum to zero. We then set $b = \prod i b_i \pmod{n}$ and $c = \prod p_j^{\frac{1}{2} \sum \alpha_{ij}}$ where p_j are the elements of $B \setminus \{-1\}$ and the sum is taken over the same subset of i . If $B \not\equiv \pm c \pmod{n}$, then $\text{g.c.d.}(b + c, n)$ is a nontrivial factor of n . If $b \equiv \pm c \pmod{n}$, then we evaluate a different subset of i such that $\sum \vec{B}_i = 0$. If this is not possible, then more a_i , b_i , and $b_i^2 \pmod{n}$ must be computed to expand our factor base B . [6]

Which Method Is The Best?: The Advantages and Disadvantages of Each Method

While there are other methods of factoring that could be used (e.g. elliptic curve factorization, index calculus method, etc.) in order to crack RSA cryptography, these three methods are the simplest factoring methods that I could find. So now that we have these three methods, which one is the best for cracking RSA cryptography?

Pollard's Rho Method vs. Fermat's Method (Part I)

Pollard's rho method of factorization is the simplest factorization algorithm of the ones discussed in this paper. Simple calculations such as evaluating a chosen function, doing modular arithmetic, and finding the greatest common divisor of two numbers are all simple to do by hand or to code into a computer program (and all of these calculations are computationally inexpensive to calculate), yet there are some disadvantages to this method. The first one is that there are many computations that need to be done to find the factorization. This increase in the number of computations can lead to errors (when calculating by hand as well as coding a computer program) and increase the time it takes to find the answer. Even if the functions are computationally inexpensive to calculate in a computer program, the number of calculations that have to be done will add up and therefore take longer to find the prime factors. Second, we may not find the factor on the first attempt which means that we would have to start over with either a new function f , a

new starting value x_0 , a new way to compute the g.c.d., or any combination of these three inputs. The third problem is that there are too many choices for inputs. It is not something that can immediately be executed [like Fermat's method (Part I)] and the choices that are made do not guarantee success, thus causing our choices to be re-evaluated for the second round of calculations if we do not get a proper factor.

Fermat's method (Part I), on the other hand, is a little more complex than the rho method in its computations, yet it is immediately executable and will give us the answer the first time (even if it takes a while). The disadvantage to Fermat's method is that when it is being executed by a computer program there is a chance that there is a problem with overflow errors (e.g. t^2 or n exceed the upper bounds of the computer's computational limits).

In order to examine each method further I wrote C programs for both methods. For each method I tested 24 numbers that were 19 digits in length (the highest my computer could test) where the primes were randomly generated and had varying distances between each other. The results for each of these numbers can be seen in Table 4 on page 15, but my findings can be summarized in Table 3 where I look at the average time each method took to calculate the prime factors as well as if the method is successful on this first try. There are two things that are important to note about the results. The first is that for the rho method I used $f(x) = x^2 + x + 1$, $x_0 = 2$, and I used the first method to calculate the g.c.d. that was outlined in the rho method section on page 6. The second thing is that for Fermat's method (Part I), I was only able to write a program using Equations 7 and 8 and not Equations 9 and 10 since I was using values of n close to the upper bound of my computer's limits and calculating kn in Equation 9 would have caused overflow errors. As Table 3 shows, the rho method took an average time of 26 minutes to compute an answer and 33% of the time it was able to find the correct prime factorization of n . This means that 66% of the time it did not give me an answer, but it could with different initial conditions. Conversely, Fermat's method took an average time of less than a tenth of a second and was always able to find the factorization of n . Therefore, when comparing these two methods of factorization, I have found that Fermat's method of factorization is faster and more reliable.

Rho Method Average Time	Rho Method Found Factors (%)	Fermat Method Average Time	Fermat Method Found Factors (%)
1583.546 sec. \approx 26.392 min.	8/24 = 1/3 = 33.333%	<.147 seconds	24/24 = 100%

Table 3: Summary of Results

Fermat's Method (Part I) vs. Continued Fraction Method and Fermat's Method (Part II)

Since the continued fraction method of factorization is a refinement of Fermat's factorization with factor bases, there is no point in comparing these two methods. Instead, we shall compare Fermat's method (Part I) and the continued fraction method. We can see from the descriptions of the Fermat's method (Part II) and continued fraction method that the process is more intricate and complex than Fermat's method (Part I), yet the mathematics involved is still elementary. While I was able to write a C program for Fermat's method (Part I), I was not able to for Fermat's method (Part II) or the continued fraction method due to the complexity of the process and my introductory knowledge of C programming. Thus I was not able to compare the two methods as I did with Pollard's rho method and Fermat's method (Part I). Assuming that I had the knowledge to write such a program, the calculations involved can be programmed on a computer with single-precision floating point and it lends itself to parallel processing. The computers can even be linked by e-mail to complete the task. The disadvantage is that it is very memory-intensive so that you need a good deal of computing power, while Fermat's method (Part I) can be done on a calculator. Also, the continued fraction method works best for numbers that are roughly 300 bits long. The reason for this is that the continued fraction method requires that we find primes less than the number we are factoring, and primality testing and trial division of large numbers becomes increasingly difficult. Therefore, for numbers that are roughly 300 bits long the continued fraction method works as well as Fermat's method (Part I) but can find the results quicker. For numbers greater than 300 bits long, it can still find the answer but Fermat's method (Part I) will probably be quicker. [2][5][8]

Conclusion of Cracking RSA Cryptography

We have looked at three different methods of factorization in this paper, and while they all have their advantages and disadvantages, they are all capable of factoring the public key in RSA

cryptography. So is RSA cryptography secure? Yes. But how can that be? The first reason why this is true is the size of the public key. When I was factoring numbers with the computer programs I wrote, the largest number my computer could handle was 9, 000, 000, 114, 000, 000, 361 which is 19 digits long and less than 64 bits. RSA keys are typically 1024-2048 bits long [9], meaning that the public key n is $2^{1024} - 1 \leq n \leq 2^{2048} - 1$. That means the public key is somewhere between 309-617 digits long which is outside the bounds of any personal computer and cannot be factored with any C programs of the factorization methods above. The second reason RSA cryptography is secure is that even if we had a computer that could factor these numbers, it would take an extremely long time. The primes chosen to create n follow a set of conditions (see page 5) such that factorization is made extremely difficult and maximizes the amount of time to find its factorization. Also, the people who use RSA cryptogrpahy switch their keys on a regular basis. So by the time we have found the factorization of their public key, they have started using a new key and our work has been for nothing (except the satisfaction that we now know the factorization of a very large number, but that is not worth much). Therefore, RSA cryptography is secure not because it is uncrackable, but the amount of time and effort that must go into to cracking it is not realistic or feasible.

A Computer Calculation Results

Number Tested	Rho Method Time (in seconds)	Prime Factors Found?	Fermat Method Time (in seconds)	Prime Factors Found?
1086193063168655033	561.631	Yes	2.293	Yes
1528306804109149607	56.908	No	0.015	Yes
1725789628106787091	279.489	Yes	<0.001	Yes
2322606847041148253	55.489	Yes	0.046	Yes
2621988641265561167	261.658	Yes	0.062	Yes
2973883149614142557	471.385	No	0.249	Yes
3062120511377001367	267.836	Yes	<0.001	Yes
3549984033641742101	1348.027	No	0.015	Yes
3990000139835001221	393.666	Yes	<0.001	Yes
4395300230629003021	1107.490	Yes	<0.001	Yes
4548400130772808207	1205.193	Yes	<0.001	Yes
4745906468019226637	3251.851	No	0.015	Yes
5485632351956318231	3033.154	No	<0.001	Yes
5759222164792545217	796.848	No	<0.001	Yes
5972673451757577503	3542.182	No	0.592	Yes
6022074550288762351	1622.238	No	<0.001	Yes
6210830378410100153	2818.655	No	<0.001	Yes
6717045154677851171	3004.887	No	<0.001	Yes
7047000270990002597	317.569	No	0.062	Yes
7450265298435973823	1616.971	No	0.015	Yes
778319555846063313	3938.641	No	<0.001	Yes
8016652681620953801	5768.512	No	0.156	Yes
8793658959019646351	2056.314	No	<0.001	Yes
8999850293998152257	228.524	No	<0.001	Yes

Table 4: Results of Computer Program for Rho and Fermat Method

B Glossary

B-number - an integer b where the least absolute residue $b^2 \pmod n$ can be written as a product of numbers in the factor base B .
ciphertext - the encrypted message.
cryptanalyst - someone who tries to break cryptographic systems without knowledge of the key.
cryptanalysis - the study of breaking cryptographic systems.
cryptographers - someone who tries to create unbeatable cryptographic systems.
cryptography - the study of methods of sending encrypted messages so that only the intended recipient can decrypt and read the message.
deciphering/decryption - process of converting ciphertext to plaintext.
enciphering/encryption - process of converting plaintext to ciphertext.
factor base - a set $B = \{p_1, p_2, \dots, p_k\}$ of distinct primes (except p_i may be -1).
key - the element that turns the general encryption algorithm into a specific method of encryption.
least absolute residue of $a \pmod n$ - is the integer in the interval $\left[\frac{-n}{2}, \frac{n}{2} \right]$ to which a is congruent.
plaintext - the original message to be sent.
private key - the private information used by the receiver to decrypt the message sent.
public key - the algorithm and public information that is used by the sender to encrypt the message.

recurrence relation - an equation that recursively defines a sequence; using a recursive definition to define an equation.
residue classes modulo n - the set of equivalence classes congruent modulo n , meaning that any integer is represented on the set $\{0, 1, \dots, n-1\}$.

C Computer Code for Pollard's Rho Method

```
#include<stdio.h>
#include<stdlib.h>
#include<time.h>

long long f(long long x, long long int n)
{
    long long ans;
    ans=(x*x+x+1)%n;
    return ans;
}

long long gcd(long long a, long long b)
{
    long long c;
    do
    {
        c=a%b;
        a=b;
        b=c;
    }while(c!=0);
    return a;
}

int main(void)
{
    unsigned long long n, x=2, y, z, w, p, q;
    int counter=0, h=1, k=0;
    printf("What is n?\n");
    scanf("%llu", &n);
    printf("\n");
    clock_t start = clock();
    do
    {
        counter++;
        if(k<counter)
            do
            {
                h=2*h;
                k=h-1;
            }while(k<counter);
        y=f(x,n);
        w=abs(y-z);
        p=gcd(w, n);
        if(k==counter)
            {
                z=y;
                x=f(x,n);
                if(p==n)
                    {
                        printf("The factors of %llu are %llu and %llu.\n", n, p, p);
                        return 0;
                    }
                else
                    {
                        printf("Time elapsed: %f seconds\n", ((double)clock() - start) / CLOCKS_PER_SEC);
                        printf("The factors of %llu are %llu and %llu.\n", n, p, p);
                        return 0;
                    }
            }
    }while(counter<1000000);
}
```

```
{
    z=y;
}
x=y;
}while(p==1);
q=n/p;
printf("Time elapsed: %f seconds\n", ((double)clock() - start) / CLOCKS_PER_SEC);
printf("The factors of %llu are %llu and %llu.\n", n, q, p);
return 0;
}
```

D Computer Code for Fermat's Factorization (Part I)

```
#include<stdio.h>
#include<math.h>
#include<ctime.h>
int main(void)
{
    unsigned long long int n, p, q, t, s1, x;
    long double n2, sroot, sroot2, s;
    int err=0;
    printf("What number do you wish to factor?n");
    scanf("%llu", &n);
    printf("%llu", n);
    clock_t start = clock();
    n2=(long double)n;
    sroot = sqrt(n2);
    sroot2 = floor(sroot);
    t = (unsigned long long int) sroot2;
    do
    {
        t++;
        x = (t*t)-n;
        s=sqrt((long double) x);
        err++;
    }while(floor(s)!=s);
    s1 = (unsigned long long int) s;
    p=t+s1;
    q=t-s1;
    printf("Time elapsed: %f seconds\n", ((double)clock() - start) / CLOCKS_PER_SEC);
    printf("The factors of %llu are %llu and %llu.\n", n, p, q);
    return 0;
}
```

E Picture Credits

Figure 1 taken from http://www.thedavincigame.com/Boardgame/Frequency_Analysis.png
Figure 2 taken from <http://enigma.wikispaces.com/file/view/Vigenere-square.png/30615996/Vigenere-square.png>
Figure 3 taken from <http://www.mathmu.cn/doc/images/rho.png>

References

- [1] G.E. Andrews, *Number Theory*, Dover Publications, Inc., New York, 1971.
- [2] R. Crandall, C. Pomerance, *Prime Numbers: A Computational Perspective*, Springer, New York, 2005.
- [3] H. Davenport, *The Higher Arithmetic: An Introduction to the Theory of Numbers*, Cambridge University Press, Cambridge, 2008.
- [4] W. Diffie, M. Hellman, New Directions in Cryptography, *IEEE Trans. Inf. Theory*, IT-22, 1976, 644-654.
- [5] J. Katz, Y. Lindell, *Introduction to Modern Cryptography*, Chapman & Hall/CRC, New York, 2008.
- [6] N. Koblitz, A Course in Number Theory and Cryptography, Springer-Verlag, New York, 1987.
- [7] N. Koblitz, A. Menezes, The Brave New World of Bodacious Assumptions in Cryptography, *Notices of the AMS*, 57(3), 2010, 357-365.
- [8] I. Niven, H.S. Zuckerman, H.L. Montgomery, *An Introduction to the Theory of Numbers*, John Wiley & Sons, Inc., New York, 1960.
- [9] RSA Cryptography available at <<http://en.wikipedia.org/wiki/RSA>>.
- [10] S. Singh, *The Code Book: The Science of Secrecy From Ancient Egypt to Quantum Cryptography*, Anchor Books, New York, 1999.

Stability Analysis of FitzHugh – Nagumo with Smooth Periodic Forcing

Tyler Massaro

1 Background

As Izhikevich so aptly put it,

“If somebody were to put a gun to the head of the author of this book and ask him to name the single most important concept in brain science, he would say it is the concept of a *neuron*[16].”

By no means are the concepts forwarded in his book restricted to brain science. Indeed, one may use the same techniques when studying most any physiological system of the human body in which neurons play an active role. Certainly this is the case for studying cardiac dynamics.

On a larger scale, neurons form an incredibly complex network that branches to innervate the entire body of an organism; it is estimated that a typical neuron communicates directly with over 10,000 other neurons [16]. This communication between neurons takes the form of the delivery and subsequent reception of a traveling electric wave, called an *action potential*[1]. These action potentials become the subject of Hodgkin and Huxley’s groundbreaking research.

At any given time, the neuron possesses a certain voltage difference across its membrane, known as its potential. To keep the membrane potential regulated, the neuron is constantly adjusting the flow of ions into and out of the cell. The movement of any ion across the membrane is detectable as an electric current. Hence, it follows, that any accumulation of ions on one side of the membrane or the other will result in a change in the membrane potential. When the membrane potential is 0, there is a balance of charges inside and outside of the membrane.

Before we begin looking at Hodgkin and Huxley’s model, we must first understand *how* the membrane adjusts the flow of ions into and out of the cell. Within the cell, there is a predominance of potassium, K^+ , ions. To keep K^+ ions inside of the cell, there are pumps located on the membrane that use energy to actively transport K^+ in but not out. Leaving the cell is actually a much easier task for K^+ : there are *leak channels* that “randomly flicker between open and closed states no matter what conditions are inside or outside the cell...when they are open, they allow K^+ to move freely[1].”

Since the concentration of K^+ ions is so much higher inside the cell than outside, there is a tendency for K^+ to flow out of these leak channels along its concentration gradient. When this happens, there is a negative charge left behind by the K^+ ions immediately leaving the cell. This build-up of negative charge is actually enough to, in a sense, catch the K^+ ions in the act of leaving and momentarily halt the flow of charge across the membrane. At this precise moment, “the electrochemical gradient of K^+ is zero, even though there is still a much higher concentration of K^+ inside of the cell than out[1].” For any cell, the *resting membrane potential* is achieved whenever the total flow of ions across the cell membrane is balanced by the charge existing inside of the cell. We may use an adapted version of the *Nernst Equation* to determine the resting membrane potential with respect to a particular ion[1]:

$$V = 62\log_{10}C_o/C_i,$$

where V is the membrane potential (in mV), C_o is the ion concentration outside of the cell, and C_i is the ion concentration inside of the cell.

Before we continue, it is important to revisit the concept of action potentials. Neurons communicate with each other through the use of electric signals which alter the membrane potential on the recipient neuron. To continue propagating this message, the change in membrane potential must travel the length of the entire cell to the next recipient. Across short distances, this is not a problem. However, longer distances prove to be a bit more of a challenge. To compensate for any amplification that may need to take place, the input of an amount of electrical stimulation beyond a certain threshold yields our aforementioned action potential. These action potentials can carry messages at speeds of up to 100 meters per second[1].

Physiologically speaking, there are some key events taking place whenever an action potential is discharged. Once the cell receives a sufficient electrical stimulus, the membrane is rapidly depolarized; that is to say, the membrane potential becomes less negative. The membrane depolarization causes voltage-gated Na^+ channels to open. (At this point, we have not yet discussed the

role of sodium in the cell. The important thing to understand is that the concentration of sodium is higher outside of the cell than on the inside.) When these Na^+ channels open up, they allow sodium ions to travel down their concentration gradient into the cell. This in turn causes more depolarization, which causes more channels to open. The end result, occurring in less than 1 millisecond, is a shift in membrane potential from its typical resting value of about -60mV to somewhere around $+40\text{mV}$ [1]. The value of $+40\text{mV}$ actually represents the resting potential for sodium, and so at this point no more sodium ions are entering the cell.

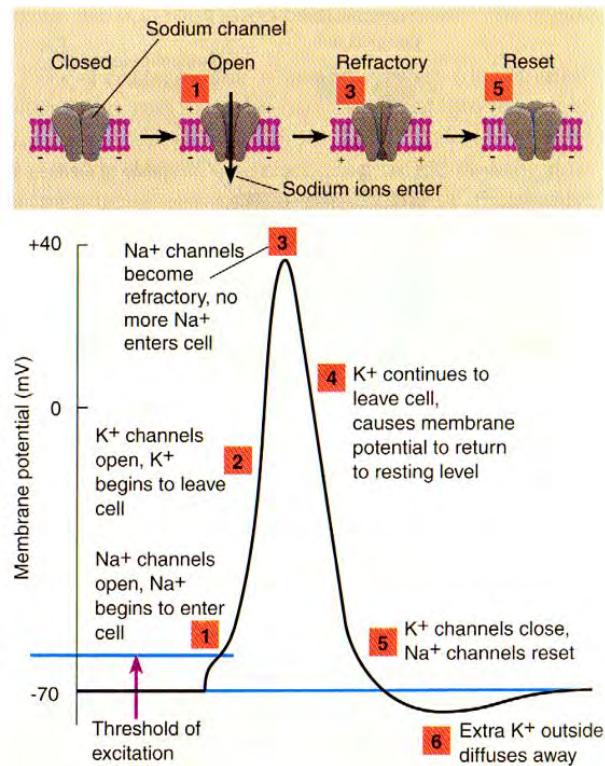


Figure 1 - Diagram of Action Potential
[http://internal.psychology.illinois.edu/~etaylor4/
action_potential.jpg](http://internal.psychology.illinois.edu/~etaylor4/action_potential.jpg)

Before the cell is ready to respond to another signal, it must first return to its resting membrane potential. This is accomplished in a couple of different ways. First, once all of the sodium channels have opened, they switch to an inactive conformation which prevents them from opening back up (imagine putting up a wall in front of an open door). Since the membrane is still depolarized at this point, the gates will stay open. This inactive conformation will persist as long as the membrane is sufficiently depolarized. Once the

membrane potential goes back down, the sodium channels switch from inactive to closed [1].

At the same time that all of this is occurring, there are also potassium channels that have been opened due to the membrane depolarization. There is a time lag which prevents the potassium gates from responding as quickly as those for sodium. However, as soon as these changes are opened, the K^+ ions are able to travel down their concentration gradient out of the cell, carrying positive charges out with them. The result is a sudden repolarization of the cell. This causes it to return to its resting membrane potential, and we start the process all over again [1].

As a special note of interest, cardiac cells are slightly different from nerve cells in that there are actually two repolarization steps taking place once the influx of sodium has sufficiently depolarized the cell: *fast repolarization* from the exit of K^+ ions, and *slow repolarization* that takes place due to an increase in Ca^{2+} conductance [26]. For now, we will continue dealing solely with Na^+ and K^+ .

At this point, it is time to take a look at the models these physiological processes inspired. Arguably the most important of these was created by Alan Lloyd Hodgkin and Andrew Huxley, two men who forever changed the landscape of mathematical biology, when in 1952, they modeled the neuronal dynamics of the squid giant axon. Below are listed the complete set of space-clamped Hodgkin-Huxley equations [14, 16]:

$$\begin{aligned} C\dot{V} &= I - \overbrace{\bar{g}_K n^4 (V - E_K)}^{I_K} - \overbrace{\bar{g}_{\text{Na}} m^3 h (V - E_{\text{Na}})}^{I_{\text{Na}}} - \overbrace{g_L (V - E_L)}^{I_L} \\ \dot{n} &= \alpha_n(V)(1-n) - \beta_n(V)n \\ \dot{m} &= \alpha_m(V)(1-m) - \beta_m(V)m \\ \dot{h} &= \alpha_h(V)(1-h) - \beta_h(V), \end{aligned}$$

where

$$\begin{aligned} \alpha_n(V) &= 0.01 \frac{10-V}{\exp(\frac{10-V}{10})-1}, \\ \beta_n(V) &= 0.125 \exp\left(\frac{-V}{80}\right), \\ \alpha_m(V) &= 0.1 \frac{25-V}{\exp(\frac{25-V}{10})-1}, \\ \beta_m(V) &= 4 \exp\left(\frac{-V}{18}\right), \\ \alpha_h(V) &= 0.07 \exp\left(\frac{-V}{20}\right), \\ \beta_h &= \frac{1}{\exp(\frac{30-V}{10})+1}. \end{aligned}$$

The E values represent shifted Nernst equilibrium potentials, C is the capacitance, and the g, g values represent maximal conductances. Our state variables are as follows: V is the membrane potential, and m, n, h are the activation variable for

the voltage-gated transient sodium current (there are three), the voltage-gated persistent potassium current (there are four), and the inactivation gate (there is only one), respectively. Each of the activation variables represents the empirically determined probability that a particular channel will be open based upon the current membrane potential.

To provide some background, let us consider two equations from physics. Our first is the standard version of *Ohm's Law*[27]:

$$V = IR,$$

where V is the total voltage of a circuit, I is the total current, and R is the total resistance. Our second equation is *Ohm's Law for Capacitors* [27]:

$$CV = I,$$

where C is the total capacitance, V is the change in voltage, and I is the instantaneous current passing through the capacitor.

Returning to the first line of the H-H equations, we see that this is simply an implementation of Ohm's Law for Capacitors, with the right-hand side showing a total summation of currents for each particular channel, plus the injected current, I . The terms of each of the currents may look somewhat unfamiliar, since they include the product of the conductance, a voltage, and a value representing the percentage of those particular channels that are open. Recall that conductance is simply the inverse of resistance, and it is apparent now that each of these terms may be derived from *Ohm's Law*. In general, we have the following equation, given particular membrane conductances (g_i 's) and reverse potentials (E_i 's) [16]:

$$\dot{CV} = I - \sum_i g_i(V - E_i).$$

Taking a look at the next three lines, we see the equations for the activation of variables. The α and β terms represent the different Boltzmann and Gaussian functions, which together describe the steady-state activation curve for each particular gating variable [16]. In other words, m , n , and h represent, respectively, the voltage-dependent probability that the sodium, potassium or inactivation gate is open.

Shortly after Hodgkin and Huxley published their model, biophysicist Richard FitzHugh began

an in-depth analysis of their work. He discovered that, while their model accurately captures the excitable behavior exhibited by neurons, it is difficult to fully understand *why* the math is in fact correct. This is due not to any oversight on the part of Hodgkin and Huxley, but rather because their model exists in four (4) dimensions. To alleviate this problem, FitzHugh proposed his own two-dimensional differential equation model. It combines a model from Bonhoeffer explaining the "behavior of passivated iron wires," as well as a generalized version of the van der Pol relaxation oscillator [14]. His equations, which he originally titled the Bonhoeffer-van der Pol (BVPD) oscillator, are shown below [14, 26]:

In his model, for which Jin-Ichi Nagumo constructed the equivalent circuit the following year in 1962, x "mimics the membrane voltage," while y represents a recovery variable, or "activation of the outward current [16]." Both a

$$\begin{cases} \dot{x} = c(y + x - x^3/3 + z), \\ \dot{y} = -(x - a + by)/c, \end{cases}$$

$$1 - 2b/3 < a < 1, \quad 0 < b < 1, \quad b < c^2.$$

and b are constants he supplied (in his 1961 paper, FitzHugh fixes $a = 0.7$ and $b = 0.8$). The third constant, c , is left over from the derivation of the BVPD oscillator (he fixes $c = 3$). The last variable, z , represents the injected current. It is important to note that in the case of $a = b = z = 0$, the model becomes the original van de Pol oscillator [14].

Many different versions of this model exist [16, 17, 26], all of them differing by some kind of transform of variables. We will consider the model used by Kostova et al. in their paper, which presents the FitzHugh-Nagumo model without diffusion:

$$\begin{cases} \frac{du}{dt} = \varepsilon g(u) - w + I, \\ \frac{dw}{dt} = u - aw, \end{cases}$$

where $g(u) = u(u - \lambda)(1 - u)$, $0 < \lambda < 1$ and $a, \varepsilon > 0$ [17]. Here, the state variable u is the voltage, w is the recovery variable, and I is the injected current.

2 Stability Analysis via a Linear Approximation

2.1 Examining the Nullclines

When studying dynamical systems, it is important to be familiar with the concept of nullclines. In a broader sense, a nullcline is simply an isocline, or a curve along which the value of a derivative is constant. In particular, the nullcline is the curve along which the value of the derivative is zero. Taking another look at FH-N (1.1), we see that there are two potential nullclines, one where the derivative of u will be zero, and the other where the derivative of w will be zero:

$$\begin{cases} \frac{du}{dt} = \varepsilon g(u) - w + I = 0, \\ \frac{dw}{dt} = u - aw = 0. \end{cases}$$

One of these nullclines is cubic, and the other is linear. Consider an intersection of those two graphs. At that particular point, we know that

$$\frac{du}{dt} = \frac{dw}{dt} = 0.$$

Hence, at this point, neither of our equations is changing. This point where our nullclines intersect is called an equilibrium, or fixed point. Since our nullclines are cubic and a line, geometrically we see that there could be as many as three possible intersections, and no fewer than one. Let us consider the case where $I = 0$. Our system becomes:

$$\begin{cases} \frac{du}{dt} = \varepsilon g(u) - w = 0, \\ \frac{dw}{dt} = u - aw = 0. \end{cases}$$

Evaluating the system at the origin, where $u = w = 0$, we can see that this is always an equilibrium when $I = 0$.

2.2 Linearizing FitzHugh - Nagumo

Unless otherwise state, we will assume $I = 0$ for the next few sections. Similarly, (u_e, w_e) will always refer to an equilibrium of FH-N (not necessarily the origin). Let us define the functions f_1 and f_2 as the following:

$$\begin{aligned} f_1 &:= \varepsilon g(u) - w, \\ f_2 &:= u - aw. \end{aligned}$$

Finally, we also set $b_1 = g'(u_e)$, a notation we get from Kostova et al. [17].

2.2.1 Creating a Jacobian

We may linearize FH-N by constructing a Jacobian Matrix as follows:

$$\begin{aligned} J(u, w) &:= \begin{bmatrix} \frac{\partial f_1}{\partial u} & \frac{\partial f_1}{\partial w} \\ \frac{\partial f_2}{\partial u} & \frac{\partial f_2}{\partial w} \end{bmatrix}, \\ J(u_e, w_e) &:= \begin{bmatrix} \varepsilon b_1 & -1 \\ 1 & -a \end{bmatrix}. \end{aligned}$$

We see that for any equilibrium, $J(u_e, w_e)$ has the same form, since we have the substitution in place for b_1 . Thus, we may generalize the eigenvalues of the above Jacobian to be the eigenvalues of any equilibrium. Solving the characteristic polynomial for our Jacobian, we get the following eigenvalues:

$$\mu_{1,2} = \frac{1}{2}(\varepsilon b_1 - a) \pm \frac{1}{2}\sqrt{(a - \varepsilon b_1)^2 + 4(a\varepsilon b_1 - 1)}.$$

As long as it is never the case that $\text{Re}(\mu_1) = \text{Re}(\mu_2) = 0$, the eigenvalues will always have a real part, and then our equilibrium is hyperbolic (see definition below). By the *Hartman - Grobman Theorem*, we know that we may use the Jacobian to analyze the stability of any fixed point of FH-N.

Hyperbolic Fixed Points (2-D). If $\text{Re}(\mu) \neq 0$ for both eigenvalues, the fixed point is *hyperbolic* [29].

The Hartman-Grobman Theorem. The local phase portrait near a hyperbolic fixed point is “topologically equivalent” to the phase portrait of the linearization; in particular, the stability type of the fixed point is faithfully captured by the linearization. Here topologically equivalent means that there is a homeomorphism that maps one local phase portrait onto the other, such that trajectories map onto trajectories and the sense of time is preserved [29].

2.2.2 Trace, Determinant, and Eigenvalues

From Poole [24], we find two results which tie together the trace, J , and determinant, Δ , of a matrix with its eigenvalues. For any $n \times n$ matrix, A , with a complete set of eigenvalues, $(\mu_1, \mu_2, \dots, \mu_n)$, we know:

$$\Delta_A = \mu_1 \mu_2 \dots \mu_n, \text{ and}$$

$$\tau_A = \mu_1 + \mu_2 + \dots + \mu_n.$$

For our Jacobian evaluated at an equilibrium, we have:

$$\Delta_J = 1 - \varepsilon b_1 a,$$

$$\tau_J = \varepsilon b_1 - a.$$

Using the table in theorem 12.2.1 (see below, Figure 2) from Nagle [23], we may find the different types of fixed points for each given pair of eigenvalues. We now explore the different stability cases for a given set of real eigenvalues.

Figure 2 – Stability of Linear Systems [23]

Theorem 1. Assume the origin $(0, 0)$ is an isolated critical point for the linear system

$$x'(t) = ax + by,$$

$$y'(t) = cx + dy,$$

where a, b, c , and d are real and $ad - bc \neq 0$. Let r_1 and r_2 be the roots of the characteristic equation

$$r^2 - (a + d)r + (ad - bc) = 0.$$

The stability of the origin and the classification of the origin as a critical point depends on the roots r_1 and r_2 as follows:

Roots	Type of Critical Point	Stability
distinct, positive	improper node	unstable
distinct, negative	improper node	asymptotically stable
opposite signs	saddle point	unstable
equal, positive	proper or improper node	unstable
equal, negative	proper or improper node	asymptotically stable
complex-valued:		
positive real part	spiral point	unstable
negative real part	spiral point	asymptotically stable
pure imaginary	center	stable

Case 1. Let $\varepsilon ab_1 < 1$. Then $\Delta_J > 0$. Evaluating the trace, we see that for $\varepsilon b_1 > a$, we get $\tau_J > 0$, which therefore means that we have a dominate positive eigenvalue. Since $\Delta_J > 0$, we know that both of our eigenvalues must then be positive. This gives us an unstable source. For $\varepsilon b_1 < a$, we get $\tau_J < 0$. This time however, since $\Delta_J > 0$, both of our eigenvalues are negative, and so the system is a stable sink.

Case 2. Let $\varepsilon ab_1 > 1$. Then $\Delta_J < 0$. Hence, our eigenvalues are different signs. In this case, the equilibrium is an unstable saddle.

2.3 Bifurcation Analysis

An important area to study in the field of dynamics is *bifurcation theory*. A bifurcation

occurs whenever a certain parameter in a system of equations is changed in a way that results in the creation or destruction of a fixed point. Although there are many different classifications of bifurcations, we will focus only on one.

2.3.1 Hopf Bifurcation

Consider the complex plane. In a 2-D system, such as FH-N, a stable equilibrium will have eigenvalues that lie in the left half of the plane, that is, the $\text{Re}(\mu) < 0$ half of the plane. Since these eigenvalues in general are the solutions to a particular quadratic equation, we need them both to be either real or negative, or complex conjugates in the same $\text{Re}(\mu) < 0$ part of the plane. Given a stable equilibrium, we may de-stabilize it by moving one or both of the eigenvalues to the $\text{Re}(\mu) > 0$ part of the complex plane. Once an equilibrium has been de-stabilized in this manner, a *Hopf bifurcation* has occurred [29].

2.3.2 Proposition 3.1 (From Kostova)

As the eigenvalues μ_1, μ_2 of any equilibrium (μ_1, μ_2) are of the form

$$\mu_{1,2} = \frac{1}{2}R(\varepsilon, a, b_1) \pm \frac{1}{2}\sqrt{R^2 + 4Q}$$

where $Q(\varepsilon, a, b_1) \equiv \varepsilon, a, b_1 - 1$ and $R \equiv \varepsilon b_1 - a$ a Hopf bifurcation occurs in cases when $R = 0$ and $Q < 0$ [17].

Proof. Recall from earlier that we defined the Jacobian for FH-N as follows:

$$J(u, w) = \begin{bmatrix} \varepsilon g'(u) & -1 \\ 1 & -a \end{bmatrix}.$$

Now we solve for the eigenvalues of this matrix evaluated at an equilibrium. For equation 2.1 we know our eigenvalues have the following form:

$$\mu_{1,2} = \frac{1}{2}(\varepsilon b_1 - a) \pm \frac{1}{2}\sqrt{(a - \varepsilon b_1)^2 - 4(-a\varepsilon b_1 + 1)}.$$

Substituting in now for R and Q , we clearly have $\mu_{1,2} = \frac{1}{2}R \pm \frac{1}{2}\sqrt{R^2 + 4Q}$.

If we allow $Q < 0$ and $R = 0$, our eigenvalues become:

$$\mu_{1,2} = \pm \frac{1}{2}\sqrt{4Q} = \pm i\sqrt{|Q|}.$$

Both of these eigenvalues are along the imaginary axis. This is the exact point at which Hopf bifurcation occurs.

3 Stability Analysis via Lyapunov's Second Method

At the end of nineteenth century, Russian mathematician Aleksandr Lyapunov developed an entirely new approach to analyzing the stability of nonlinear systems. His technique, now referred to as the direct method, yields more insights about equilibrium stability than the comparable linear approach. Specifically, in addition to standard classification of fixed points, one may also assess regions of asymptotic stability. However, more thorough results come at the cost of requiring the use of special auxiliary functions [6]. The ability to create these functions has been described as more of an art than an actual science, with one author going so far as to suggest that “[d]ivine intervention is usually required... [29].” In this next section, we analyze the Lyapunov functional published by Kostova et al. in their paper [17].

3.1 Defining Positive and Negative Definite

Primary among these auxiliary functions, and described in terms of conservation laws, are the total energy of a system, V , and its derivative, which Brauer and Nohel refer to as V' . Keeping these in mind, it is important to become familiar with the definitions described below for a function which is positive or negative definite in a region centered about the origin. Later on, we will consider regions of positive or negative definiteness centered about an equilibrium of FH-N (1.1), by shifting this equilibrium to the origin.

Positive Definite. The scalar function $V\left(\begin{matrix} \rightarrow \\ y \end{matrix}\right)$ is said to be *positive definite* on the set Ω if and only if $V\left(\begin{matrix} \rightarrow \\ 0 \end{matrix}\right) = 0$ and $V\left(\begin{matrix} \rightarrow \\ y \end{matrix}\right) > 0$ for $y \neq 0$ and $y \in \Omega$ [6].

Negative Definite. The scalar function $V\left(\begin{matrix} \rightarrow \\ y \end{matrix}\right)$ is *negative definite* on the set Ω if and only if $-dV\left(\begin{matrix} \rightarrow \\ y \end{matrix}\right)$ is positive definite on Ω [6].

Derivative of V . The derivative of V with respect to the system $\begin{matrix} \rightarrow \\ y^1 \end{matrix} = f\left(\begin{matrix} \rightarrow \\ y \end{matrix}\right)$ is the scalar product $V^*\left(\begin{matrix} \rightarrow \\ y \end{matrix}\right) = \text{grad } V\left(\begin{matrix} \rightarrow \\ y \end{matrix}\right) \cdot f\left(\begin{matrix} \rightarrow \\ y \end{matrix}\right)$

$$= \frac{\partial V}{\partial y_1}(\vec{y})f_1(\vec{y}) + \frac{\partial V}{\partial y_2}(\vec{y})f_2(\vec{y}) + \dots + \frac{\partial V}{\partial y_n}(\vec{y})f_n(\vec{y}) \quad [6]$$

In terms of the FitzHugh-Nagumo model, Kostova et al. offer the following for $V(u, w)$

$$V(u, w) = \frac{1}{2}[u - u_e - a(w - w_e)]^2 + G(w - w_e),$$

where $G(x) = \frac{1}{4}\epsilon ax^2 - \frac{4}{3}axb_2 - 2(b_1 - \frac{1}{a\epsilon})$. By definition, we may also find $V^*(u, w)$, $\dot{V}(u, w)$.

Before we continue, the direct method requires that our function, $V(u, w)$, be positive definite. Suppose we have some region, Ω , of the u, w -plane, which encloses an equilibrium, (u_e, w_e) . Evaluating $V(u_e, w_e)$, we get:

$$\begin{aligned} V(u_e, w_e) &= \frac{1}{2}[u_e - u_e - a(w_e - w_e)]^2 + G(w_e - w_e) \\ &= 0. \end{aligned}$$

This however is not enough to guarantee that V will be positive definite. We must now verify that $V(u, w) > 0 \forall (u, w) \in \Omega \setminus (u_e, w_e)$. To do this, we will need some notation from Kostova:

$$T = (1 - \epsilon b_1 a) - \frac{2\epsilon ab_2^2}{9}$$

and

$$S = \frac{b_2^2}{3} + b_1 - \frac{a}{\epsilon}$$

where $b_1 = g'(u_e)$ and $b_2 = \frac{g''(u_e)}{2}$ (refer to the line immediately following equation 1.1 for a definition of $g(u)$).

Let line L be defined by $L = \{(u, w) | u = u_e + aw - we\}$.

3.1.1 Proposition 2.2a (from Kostova)

$V(u, w) > 0$ for all $(u, w) \neq (u_e, w_e)$ if and only if $T > 0$. If $T \leq 0$, then $V \leq 0$ is a bounded set B , which is symmetric about the line L [17].

Proof. Consider a Taylor series expansion of $g(u)$ at the equilibrium, taking note that the terms of order 4 or greater vanish:

$$g(u) = g(u_e) + g'(u_e)(u - u_e) + \frac{1}{2}g''(u_e)(u - u_e)^2 + \frac{1}{6}g^{(3)}(u_e)(u - u_e)^3.$$

We see that $g^{(3)}(u) \equiv -6 \forall u \in \mathbb{R}$. Making this and two other substitutions for b_1 and b_2 , we get:

$$g(u) = g(u_e) + b_1(u - u_e) + b_2(u - u_e)^2 - (u - u_e)^3.$$

Before we continue, we first go back to FH-N. By definition, we know that $\frac{du}{dt} = \frac{dw}{dt} = 0$ at the equilibrium. Evaluating the system at the equilibrium, we get the following two results:

$$\begin{cases} \epsilon g(u_e) = w_e \\ u_e = aw_e. \end{cases}$$

At this point, recall how the definitions for positive and negative definiteness provided at the beginning of the section require $V(u, w)$ to be 0 at the origin of

the system. Thus far, we have only been considering regions that surround some general equilibrium. However, from section 2.1, we know that the origin will always be an equilibrium whenever $I = 0$. By introducing these transformations from Kostova, we may easily translate all of our functions to be situated around the origin:

$$v = u - u_e, \quad s = w - w_e, \quad v - as = y, \quad s = x,$$

The original FH-N system (1.1) becomes:

$$\begin{cases} \frac{dx}{dt} = y \\ \frac{dy}{dt} = -yf(x, y) - g_1(x), \end{cases} \quad (3.1)$$

where

$$\begin{aligned} f(x, y) &= \epsilon(y^2 + (3ax - b_2)y + (3a^2x^2 - 2b_2ax - b_1)) + a, \\ g_1(x) &= -\epsilon(b_1ax + b_2a^2x^2 - a^3x^3) + x. \end{aligned}$$

The line L is now described by the equation $y = 0$. Note that V and G have also changed:

$$V(x, y) = \frac{1}{2}y^2 + G(x),$$

$$G(x) = \int_0^x g_1(\xi) d\xi.$$

At this point, there are well-documented methods for determining Lyapunov functionals in a system of the form seen above in equation 3.1 [4].

Consider $V(x, y)$. We know that

$\frac{1}{2}y^2 \geq 0 \forall y \in \mathbb{R}$. To better understand $V(x, y)$ we must focus our attention on $G(x)$.

Recall that

$$G(x) = \frac{1}{4}\epsilon ax^2 \left[a^2x^2 - \frac{4}{3}axb_2 - 2\left(b_1 - \frac{1}{a\epsilon}\right)\right].$$

Already we know that $\frac{1}{4}\epsilon ax^2 \left[a^2x^2 - \frac{4}{3}axb_2 - 2\left(b_1 - \frac{1}{a\epsilon}\right)\right]$. Now, we must consider the other term. By completing the square, we see that

$$a^2x^2 - \frac{4}{3}axb_2 - 2\left(b_1 - \frac{1}{a\epsilon}\right)$$

becomes

$$\left(x - \frac{2}{3a}b_2\right)^2 - \frac{2}{a^2}\left(b_1 - \frac{1}{a\epsilon}\right) - \frac{4}{9a^2}b_2^2.$$

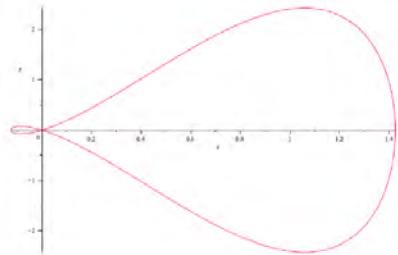
After some algebra, we introduce the constant T as described at the beginning of our proof, giving us:

$$V(x, y) = \frac{1}{2}y^2 + \frac{1}{4}\epsilon ax^2 \left[\left(x - \frac{2}{3a}b_2\right)^2 + \frac{2}{a^3\epsilon}T \right].$$

Case 1. Let $T \geq 0$. Then, $V(x, y) > 0 \forall (x, y) \neq (0, 0)$.

Case 2. Let $T < 0$. Then the level curves of V are closed curves, symmetric about the x -axis. These level curves enclose the region which we have previously described as \mathcal{B} . Shown below is the zero level curve of $V(x, y)$.

Figure 3 – $V(x, y) = 0; I = 0, a = 1.2, \lambda = 0.1, \epsilon = 14, (u_e, w_e) = (0, 0)$.



Up to this point, we have determined regions in which V is positive or negative definite. Before we may begin assessing the stability of FH-N, we must find similar regions of positive and negative definiteness for V .

3.1.2 Proposition 2.2b (from Kostova)

On L the derivative $V = \frac{\partial V}{\partial u}u + \frac{\partial V}{\partial w}w = 0$.

Additionally, $V < 0$ if $S < 0$ and $(u, w) \notin L$. If $S \geq 0$ there exists an ellipse ∂E , surrounding a region E such that : i) $V < 0$ if (u, w) belongs to the complement of $\partial E \cup E \cup L$; ii) $V > 0$ if $(u, w) \in E \setminus L$. [17].

Proof. Let us first make a change of variables to V . by definition, we have that:

$$\dot{V}(x, y) \equiv \frac{\partial V}{\partial x}\dot{x} + \frac{\partial V}{\partial y}\dot{y}.$$

Borrowing from calculations performed during the proof in Section 3.1.1, we get:

$$\dot{V}(x, y) \equiv G'(x)y + y(-yf(x, y) - g_1(x))$$

$$\begin{aligned} &\equiv \left[\frac{d}{dx} \int_0^x g_1(\xi) d\xi \right] y + y(-yf(x, y) - g_1(x)) \\ &\equiv -y^2 f(x, y). \end{aligned}$$

Recall also that the line L has the equation $y = 0$. It is then clear that $V(x, y) = 0$ on L , which includes the origin. To make further insights about $V(x, y)$ however, we must consider $f(x, y)$. From before, we have:

$$f(x, y) = \epsilon(y^2 + (3ax - b_2)y + (3a^2x^2 - 2b_2ax - b_1)) + a.$$

Rearranging terms, we get:

$$\frac{1}{\epsilon}f(x, y) = \left(y^2 + (3ax - b_2)y + \left(3a^2x^2 - 2b_2ax - b_1 + \frac{a}{\epsilon}\right)\right).$$

Completing the square for a quadratic with respect to y , we get

$$\frac{1}{\varepsilon}f(x, y) = \left(y + \left(\frac{3ax - b_2}{2}\right)\right)^2 + \left(3a^2 - \frac{9}{4}a^2\right)x^2 + \left(-2b_2a + \frac{6}{4}b_2a\right)x - \frac{1}{4}(9a^2x^2 - 6b_2ax + b_2^2).$$

Once again, we complete the square, this time using constant terms with respect to y :

$$\begin{aligned}\frac{1}{\varepsilon}f(x, y) &= \left(y + \left(\frac{3ax - b_2}{2}\right)\right)^2 + \left(3a^2 - \frac{9}{4}a^2\right)x^2 + \left(-2b_2a + \frac{6}{4}b_2a\right)x \\ &\quad + \left(-b_1 + \frac{a}{\varepsilon} - \frac{1}{4}b_2^2\right) \\ &\vdots \\ &= \left(y + \left(\frac{3ax - b_2}{2}\right)\right)^2 + \left(x - \frac{b_2}{3a}\right)^2 - \frac{4}{3a^2}\left(\frac{b_2^2}{3} + b_1 - \frac{a}{\varepsilon}\right).\end{aligned}$$

Now we substitute back in for S , seeing that our expression becomes

$$\frac{1}{\varepsilon}f(x, y) = \left(y + \left(\frac{3ax - b_2}{2}\right)\right)^2 + \left(x + \frac{b_2}{3a}\right)^2 - \frac{4}{3a^2}S.$$

In terms of \dot{V} , we now have:

$$\dot{V}(x, y) = -\varepsilon y^2 \left[\left(y + \left(\frac{3ax - b_2}{2}\right)\right)^2 + \left(x + \frac{b_2}{3a}\right)^2 - \frac{4}{3a^2}S \right]$$

Case 1. Let $S < 0$. Then $f(x, y) > 0$, and so $\dot{V}(x, y) < 0 \forall (x, y) \notin L$.

Case 2. Let $S > 0$. Then, for $f(x, y) = 0$, we have an ellipse, ∂E , surrounding a region E :

$$f(x, y) = \left(y + \left(\frac{3ax - b_2}{2}\right)\right)^2 + \left(x + \frac{b_2}{3a}\right)^2 = \frac{4}{3a^2}S.$$

Since $y = 0$ along L , we know the following:

$$\dot{V}(x, y) > 0 \forall (x, y) : (x, y) \in E \setminus (\partial E \cup L).$$

3.2 Lyapunov's Theorems for Stability

We have successfully found regions in the x , y -plane (a transformed version of the u , w -plane) where our function $V(x, y)$ and its derivative $\dot{V}(x, y)$ are positive or negative definite. The following theorems from Brauer and Nohel provide for us a way to analyze the stability of FH-N in light of the regions we have found.

Theorem 1. If there exists a scalar function $V(\vec{y})$ that is positive definite and for which $V^*(\vec{y}) \leq 0$ on some region Ω containing the origin, then the zero solution of $\vec{y}' = f(\vec{y})$ is stable [6].

Theorem 2. If there exists a scalar function $V(\vec{y})$ that is positive definite and for which $V^*(\vec{y})$ is negative definite on some region Ω containing the origin, then the zero solution of $\vec{y}' = f(\vec{y})$ is asymptotically stable [6].

Theorem 3. If there exists a scalar function $V(\vec{y})$, $V(0) = 0$, such that $V^*(\vec{y})$ is either positive definite or negative definite on some region Ω containing the origin and if there exists in every neighborhood N of the origin, $N \subset \Omega$, at least one point $a \neq 0$ such that $V(a)$ has the same sign as $V^*(\vec{y})$, then the zero solution of $\vec{y}' = f(\vec{y})$ is unstable [6].

3.2.1 Proposition 2.2c (from Kostova)

If $\varepsilon b_1 a < 1$ and $\varepsilon b_1 (u_e, w_e)$ which no solution curves of FH-N (3.1) enter. If $\varepsilon b_1 a < 1$ and $\varepsilon b_1 < a$, there is a neighborhood of the equilibrium which no solution curve leaves. These neighborhoods can be found explicitly by using level curves of V [17]

Proof. Recall that:

$$V(x, y) = \frac{1}{2}y^2 + \frac{1}{4}\varepsilon ax^2 \left| \left(x - \frac{2b_2}{3a}\right)^2 + \frac{2}{a^3\varepsilon}T \right|,$$

$$\text{where } T = (1 - \varepsilon b_1 a) - \frac{2\varepsilon ab_2^2}{9}.$$

Case 1. Let $\varepsilon b_1 a < 1$. Whenever $1 - \varepsilon b_1 a > \frac{2}{9}\varepsilon ab_2^2$, there is some neighborhood of $(u_e, w_e) = (0, 0)$ such that $V(x, y) > 0$ for all x and y in that neighborhood. Recall also that before, we defined the set, B , such that $V(x, y) \leq 0 \forall (x, y) \in B$. Hence if B exists in this case then it does not contain the origin. Consider how $\varepsilon b_1 > a$. Then, we know that S (or $\frac{b_2^2}{2} + b_1 - \frac{a}{\varepsilon}$), is strictly greater than 0. Hence, $V(x, y) > 0 \forall (x, y) \in E \setminus L$.

Evaluating $f(x, y)$ at the origin, we get:

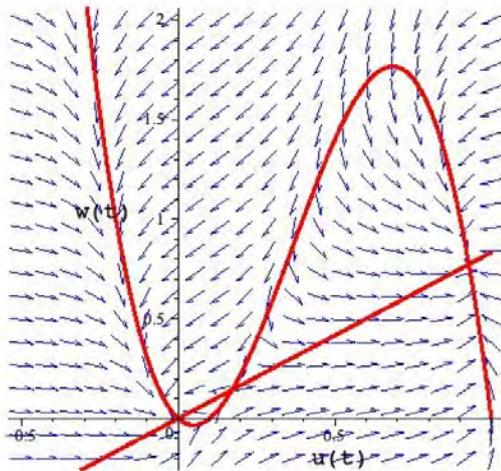
$$\begin{aligned}f(0, 0) &= \varepsilon(-b_1) + a \\ &= a - \varepsilon b_1.\end{aligned}$$

But, we have assumed $\varepsilon b_1 > a$. Then, $f(0, 0) < 0$. And so, for a neighborhood of the origin, we know that $V(x, y) > 0$, except on the line L , in which case, $V(x, 0) \equiv 0$. And by definition, this region, which contains the origin, must be part of E .

For level curves $V(x, y) = c$, $c > 0$, we allow that c surrounds B . At the same time however, we restrict this level curve to be contained entirely within E . In this way, there is a region where $V(x, y) > 0$ and $V(x, y) > 0$. Solutions may exist entirely within B , at the origin, or outside of E entirely, since E is the region where $V(x, y) > 0$.

Case 2. Let $\varepsilon b_1 < a$. Then, if $S < 0$, either no ellipse E exists, or the ellipse does exist, however the origin is no longer inside of it. Consider once again $f(x, y)$ at the origin. We see that $f(0, 0) > 0$. Hence, $V(x, y) < 0$. Recall that our ellipse, E , contains the region where $V(x, y) > 0$, assuming we restrict $(x, y) \in E$ ($E \cap L$). Our solution curve, $V(x, y) = c$ may exist anywhere so long as it does not intersect B or E .

Figure 4 – Direction Field for FitzHugh-Nagumo



4 Chaos

4.1 Butterflies

We have really only focused on determining the stability of our fixed points, however there are many other interesting questions we can ask of a dynamical system. Two of these questions, which concern sensitivity dependence, we can lump together: how sensitive is our system to the initial conditions that we give it, and how sensitive is our system to a certain parameter which it calls?

The relevance of this first question was explored by meteorologist Edward Lorenz [20]. At the time, he was studying weather forecasting models. He found that by changing his initial input to the system, he could wildly, and quite unexpectedly, change the prediction given by his model. Consider the following question, which was actually the title of a talk given by Lorenz back in 1972 [20]:

Does the Flap of a Butterfly's Wings in Brazil Set Off a Tornado in Texas?

This may at first seem frivolous, but the concept which drove him to ask in the first place digs a little bit deeper. Given some system which you use to make predictions (in essence, any mathematical model), do you expect that using roughly equal initial conditions will give you roughly the same prediction? Surprisingly, and this is what Lorenz discovered, the answer is not always yes.

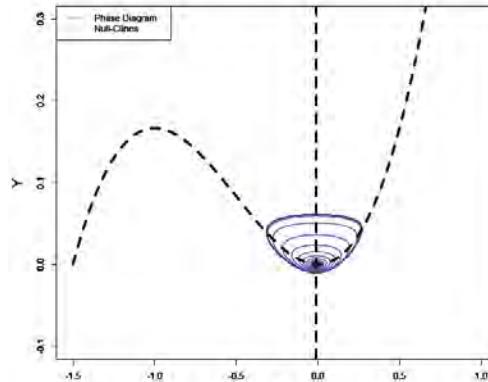
Granted, this question depends on a lot of things, for instance how far apart your initial conditions are, how far into the future you wish to make predictions, and how different predictions

need to be before you are willing to actually deem them “different.” However, once we define explicitly what we are asking, we can learn a great deal about our system. When we start thinking about this in mathematical terms, the butterfly effect means that two solutions, initialized ever so slightly apart, will diverge exponentially as time progresses (assuming of course that our system in question possesses this property).

4.2 Modified BVP with Smooth Periodic Forcing

With regards to the FitzHugh-Nagumo model, asking such a question as to whether it is sensitive to initial conditions is in most cases trivial. If we take a look at the vector field in the phase plane (see above), we see that none of our solutions will run away on some different path, since they are all restricted. You may recall at this point, back in section 3, how we found a region in which V was negative definite.

Figure 5 – Modified BVP Phase Portrait, $\kappa = 0$.



Even more specifically however, we know that each solution starting in a certain neighborhood of the equilibrium will either converge asymptotically to the equilibrium, or it will periodically trace an orbit that is held within the neighborhood. There are no surprises here: as long as you initialize a solution in the neighborhood, you will get asymptotic convergence or an orbit.

But what happens when you start changing the parameters inside of the equations themselves? We will begin to examine this question by considering a modified version of the Bonhoeffer - van der Pol equation [5], which is a distant cousin of the FitzHugh-Nagumo model (remove the forcing function and do a change of variables to get FH-N):

$$\begin{cases} \epsilon \frac{dx}{dt} = y - \left(\frac{1}{2}x^2 + \frac{1}{3}x^3 \right), & 0 < \epsilon \ll 1, \\ \frac{dy}{dt} = -(x + \alpha) + s(t), & \alpha \in \mathbb{R}. \end{cases}$$

The authors, Braaksma et al., define $s(t)$ to be a Dirac δ -function calling t modulus some constant, T . While the Dirac function is especially useful for modeling neuronal dynamics, we decided to look at smooth forcing. The function we ultimately ended up choosing is rather simple: We consider a smooth, periodic force, generated by $s(t) = \kappa \cos(t)$.

Consider the modified BVDP oscillator which fixes $\varepsilon = \alpha = 0.01$, and $\kappa = 0$. The phase diagram for a solution starting near the origin is shown in Figure 5. We will take some liberties by assuming that the physiological analog for this solution is similar to that of our original FH-N oscillator. Notice above in Figure 6 how a neuron in the active state could be modeled by a solution, $\Phi(u, w)$, which travels sufficiently far towards the “left” side of the phase space (u decreasing) before travelling up the left knee of the cubic nullcline [14]. Taking a look at the phase portrait for the above conditions (Figure 5), we see that this particular “neuron” never reaches the active state.

Keeping ε and α fixed at their value of 0.01, we now set $\kappa = 0.5$ (Figure 7). In essence, we are delivering a continuously oscillating current of electricity, the magnitude of which does not exceed 0.5. We see now that a solution with the exact same starting conditions now sweeps all the way to the left side of the space before traveling up the left knee. Referencing once again the picture of the phase space for Figure 7, we see that this solution simulates a neuron experiencing an active state (in addition to other states).

Another important aspect of this portrait worth noting is the existence of what appear to be four periodic limit cycles through which our solution travels. Shown in Figure 8 is the *bifurcation diagram* for our bifurcating parameter, κ . We see that as the value of κ changes from 0.1 to 1, solutions exist possessing 2, 3, and 4 distinct limit cycles (we see that it is consistent with the phase portrait for $\kappa = 0.5$). For κ between 0 and 0.1 however, it is unclear what is happening. It appears as though dozens of limit cycles may potential exist. Our system seems to be highly sensitive to the value of κ . The question now becomes whether or not this parameter sensitivity means that chaos is actually present.

Figure 6 – Physiological analogs for solutions to FitzHugh’s BVP model [14].

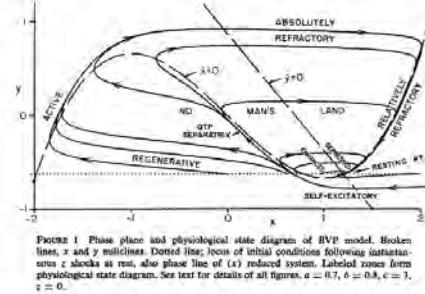
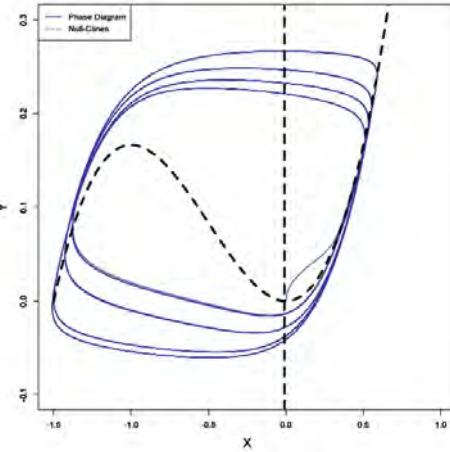


Figure 6 – Physiological analogs for solutions to FitzHugh’s BVP model [14].

Figure 7 – Modified BVDP Phase Portrait, $\kappa = 0.5$.



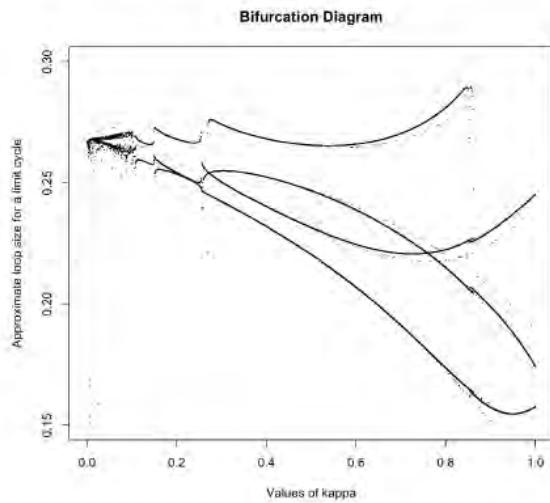
4.3 Lyapunov Exponents

Arguably the most popular way to quantify the existence of chaos is by calculating a *Lyapunov exponent*. An n -dimensional system will have n Lyapunov exponents, each corresponding to the rate of exponential divergence (or convergence) of two nearby solutions in a particular direction of the n -space. A positive value for a Lyapunov exponent indicates exponential divergence; thus, the presence of any one positive Lyapunov exponent means that the system is chaotic [34].

4.3.1 Lyapunov Spectrum Generation

There have been numerous algorithms published outlining different ways for generating what are known as *Lyapunov spectra*. As previously mentioned, an n -dimensional system will have n Lyapunov exponents. Each Lyapunov exponent is defined as the limit of the corresponding Lapunov spectrum calculated using one of these aforementioned algorithms. For our calculations, we consider the following method from Rangarajan which eliminates the need for reorthogonalization and rescaling [25].

Figure 8 – Bifurcation Diagram for κ .



Suppose we have a two dimensional system of nonlinear differential equations, like the one below:

$$\begin{cases} \frac{dx_1}{dt} = f_1(x_1, x_2), \\ \frac{dx_2}{dt} = f_2(x_1, x_2). \end{cases}$$

We may describe a Jacobian for this system in the same way as we did back in Section 2:

$$J(x_1, x_2) := \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{bmatrix}$$

Given our two dimensional system and its corresponding linearization, Rangarajan introduces three more differential equations to be coupled with the original system. The state variables λ_1 and λ_2 are the Lyapunov exponents, and θ is a third variable describing angular evolution of the solutions. The heart of the algorithm, equations for setting up the three new variables, is shown below [25]:

$$\begin{cases} \frac{d\lambda_1}{dt} = J_{11} \cos^2(\theta) + J_{22} \sin^2(\theta) - \frac{1}{2}(J_{12} + J_{21}) \sin(2\theta), \\ \frac{d\lambda_2}{dt} = J_{11} \sin^2(\theta) + J_{22} \cos^2(\theta) + \frac{1}{2}(J_{12} + J_{21}) \sin(2\theta), \\ \frac{d\theta}{dt} = -\frac{1}{2}(J_{11} - J_{22}) \sin(2\theta) + J_{12} \sin^2(\theta) - J_{21} \cos^2(\theta). \end{cases}$$

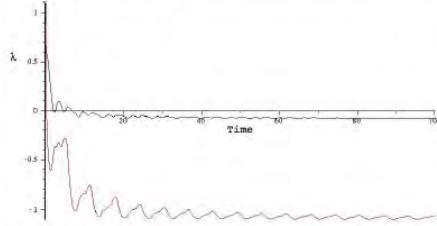
Coupling these three equations with our original system, we get a five dimensional system of differential equations. We now simultaneously solve all of these as we would any other system of differential equations, and the output corresponding

to the values of λ_1 and λ_2 over time is the Lyapunov spectrum we seek.

4.3.2 The Lyapunov Spectra

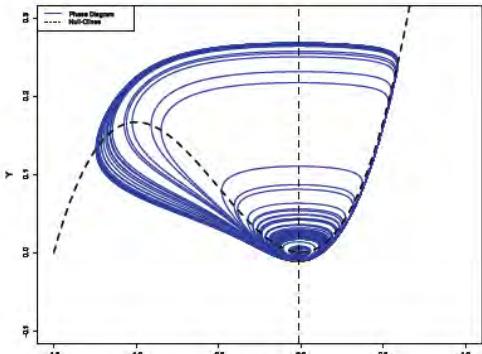
Running the algorithm for our modified BVD model with $6 = 0.5$ will produce the spectrum shown in Figure 9. Recall how we saw four stable limit cycles existing for the solution to this system. Hence, we would not expect either of our Lyapunov exponents to be greater than zero. Upon generating each of the Lyapunov spectra, we see that this is indeed the case. Both of the Lyapunov exponents for this particular system seem to settle down right away at two negative values, a result which is consistent with our expectations. In general, for roughly any system constructed with a 6 value between 0.1 and 1, we can predict, at the very least, that both of our Lyapunov exponents will be less than zero.

Figure 9 – Lyapunov Spectrum for Modified BVD, $\kappa = 0.5$.



However, the same cannot be said for systems calling a value of 6 between 0 and 0.1. Setting $6 = 0.01$, we may generate the following phase portrait (see Figure 10). Notice there are now numerous orbits, none of which are generating an active state, and none of which seem to have been traced more than once. Said another way, this solution, upon first glance at least, appears to be aperiodic. Aperiodicity is our first clue that chaos might be present in the model.

Figure 10 – Modified BVD Phase Portrait, $\kappa = 0.01$.



Changing nothing except for the value of 6, we may now generate the Lyapunov spectrum

corresponding to this new system (Figures 11 and 12). We see that one of these lines eventually makes its way underneath the horizontal axis, but the other one hovers enticingly close to the axis. At first glance, it is difficult to tell whether or not it ever actually reaches the horizontal axis and/or goes negative. The figure shown below this current spectrum zooms in on values between $t = 80$ and $t = 100$.

Figure 11 – Lyapunov Spectrum for Modified BVDP, $\kappa = 0.01$.

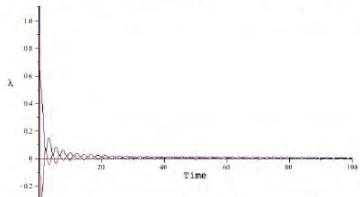
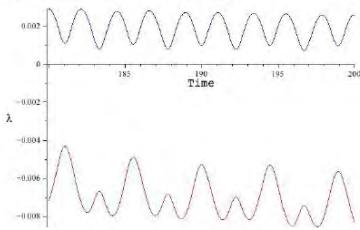


Figure 12 – Lyapunov Spectrum for Modified BVDP, $\kappa = 0.01, 180 \leq t \leq 200$.



In terms of chaos, it is difficult to judge what is happening. While one of these lines ventures below the horizontal axis, the other is clearly oscillating strictly *above* the axis. We could be remiss to immediately conclude that chaos is in fact present. And we have two reasons for offering this conjecture:

1. The oscillations are being only slightly damped, and
2. There appears to be a decreasing trend to these oscillations, suggesting they may eventually pass beneath the horizontal axis.

The first reason listed above presents issues for us since we need this output to approach some kind of limit. If it continues to behave like it is currently, we cannot say definitively whether it will asymptotically reach a limit or not (recall how the limit of $\text{cost}(t)$ is undefined as t approaches infinity). Should it *not* asymptotically approach a limit, the only real conclusion we could offer is that we need to use a more robust algorithm. The second reason is not so much a problem as it is an observation that this output could be asymptotically approaching a positive, negative, or zero valued limit. For now, all we know is that one of our

Lyapunov exponents appears to be negative, and the other is positive as far as our solver can tell us.

5 Discussion

“The healthy heart dances, while the dying organ can merely march [8]”

-Dr. Ary Goldberger, Harvard Medical School

The very nature of cardiac muscle stimulation fosters an environment for the propagation of chaos as we have previously described it. This may at first seem slightly counterintuitive. The word “chaos” itself connotes disorder. Certainly it would not immediately come to mind to describe a process as efficient as cardiac muscle contraction. And yet, what we find physiologically with heart rhythms is that a “...perfectly regular heart rhythm is actually a sign of potentially serious pathologies [10].” In particular, many periodic processes manifest themselves as arrhythmia such as ventricular fibrillation or asystole (the absence of any heartbeat whatsoever) [12]. Neither of these particular heart rhythms is conducive for sustaining life: automated external defibrillators (AEDs) were developed to counteract the presence of ventricular fibrillation in a patient; and asystole is the exact opposite of what is conducive for keeping a human alive.

At this point, it would appear as if chaos, at least in humans, is required for survival. Indeed, Harvard researcher Dr. Ary Goldberger was so moved by this idea that he made the above comment before a conference of his peers back in 1989. As the next few years unfold, it will be interesting to see what role, if any, chaos plays in assisting engineers with the development of new equipment to alter life-threatening cardiac arrhythmia in patients. The past twenty years especially have seen a tremendous increase in the demand for AEDs in public fora. Unfortunately, commercially available AEDs can only treat ventricular fibrillation and ventricular tachycardia [28].

AEDs operate by applying a burst of electricity along the natural circuitry in the heart. This electrical stimulus causes a massive depolarization event to take place, triggering simultaneous contraction of a vast majority of cardiac cells. The hope is that this sufficiently resets the heart enough for the pacemaker to regain control. In terms of a forcing function, this is almost similar to stimulation via a Dirac δ -function. Hence, we find the underlying motivation for our exploration into alternative forcing functions.

If we consider our modified BVDP model to be a sufficient analog to cardiac action potential generation, then the solution in Figure 5 roughly represents a heart experiencing ventricular fibrillation. Application of our forcing function $s(t) = 6\cos(t)$ for amplitudes between 0.1 and 1 seems to positively impact this model by inducing active states. However, it is unknown whether or not this is a realistic or even adequate portrayal of positively intervening on an arrhythmic event.

In light of the quote from Dr. Goldberger, is it possible that we should be discounting periodic solutions? If a healthy heart rhythm is in fact chaotic, would this necessitate the generation of a chaotic solution? Thus far, the closest we have come to the aforementioned chaotic solution is one which nondiscriminantly oscillates along subthreshold or superthreshold orbits, most of which do not even come close to simulating an active event in the cell. In essence, this would imply that the heart is “skipping a beat” each time it fails to generate an action potential. This is no closer to offering a viable heart rhythm, and is actually further off the mark, than our periodic solutions. Unfortunately, our search continues for an induced current that can generate both chaos and muscle contraction.

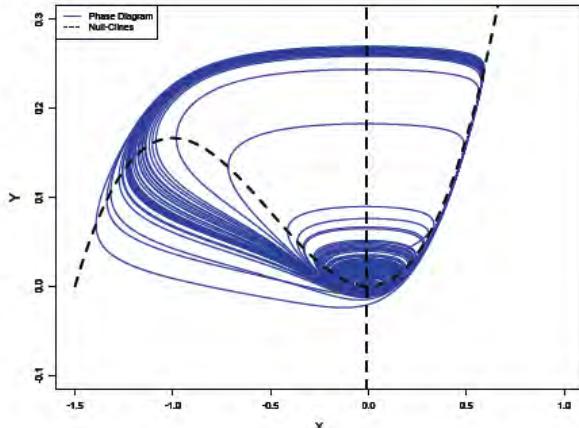
Another issue needing to be considered is the fact that we cannot, in our modified BVDP model with smooth periodic forcing, remove the forcing lest the neuron quit generating action potentials. Shown below in Figure 13 is the phase portrait for the modified BVDP model with a damped periodic forcing function, $s(t) = \frac{1}{t+1}k\cos(t)$. We see maybe one action potential generated, and then rest are all subthreshold excitations.

At first glance, it would appear as though we have to continuously induce our current. This imposes an entirely impractical, even dangerous, requirement on emergency service providers in the field. However, if our forcing function behaves at all like an AED, this result is not surprising. Once you strip away the forcing function, or in our case, once you evaluate solutions after t has grown sufficiently large, the underlying model describes a v-fib-like event taking place. It would then only make sense that action potentials are no longer generated.

The question now is whether or not our forcing function could effectively take the place of a strong induced electrical spike, similar to that delivered by an AED. And if the answer is no, are there scenarios in which continuous application of

our periodic current would be practical? Certainly no such scenario is imaginable for AEDs, however the possibility remains that it could be useful within a highly controlled setting such as inside of an operating room during surgery or built into an implantable pacemaker. Ultimately, this is a question best left to the engineers and surgeons.

Figure 13 – Modified BVDP Phase Portrait, Damped Forcing



The reason why this is so important is because sudden cardiac arrest (SCA), causes the deaths of more than 250,000 Americans each year [15]. Contrary to popular belief, SCA is first and foremost an electrical problem, triggered by faulty heart rhythms. It should not be confused with a heart attack, which is actually a blockage in one of the major blood vessels of the circulatory system. Certainly a heart attack could eventually become cardiac arrest if left untreated, but qualitatively they are entirely different events.

Whereas heart blockages and similar “plumbing problems” can be remedied by angioplasty or bypass surgery, SCA requires immediate intervention. Typically the window for successful interruption of a cardiac arrest episode will close within approximately eight to ten minutes of onset. Even with proper training, like a CPR or First Aid course that incorporates the use of an AED, SCA results in death for most out-of-hospital patients. This is certainly not for lack of trying, there are just two big problems victims currently face:

- CPR is an inefficient substitute for the natural blood delivery to the heart, and
- AEDs are only effective against two arrhythmia, v-fib and v-tach.

Ideally, technology will be made widely available so that *any* arrhythmia could be treated in an out-of-hospital environment by a layperson.

Our research has not discovered the technology described above. However, it is a step in the right direction. It is my most sincere belief that such technology can exist, and I suspect we will see it in the near future as the research progresses. In the meantime, I hope that our journey will prove useful for those looking to advance the areas of electrocardiography and AED engineering.

References

- [1] ALBERTS, B. *Essential Cell Biology, 3rd Ed.* Garland Science, New York, 2010.
- [2] ARMBRUSTER, D. The “almost” complete dynamics of the Fitzhugh-Nagumo equations. *World Scientific* (1997), 89 – 102.
- [3] AXLER, S. *Linear Algebra Done Right*, Second ed. Springer Science + Business Media, LLC, New York, 1997.
- [4] BAKER, JOHN W. Stability Properties of a Second Order Damped and Forced Nonlinear Differential Equation. *SIAM Journal of Applied Mathematics* 27, 1 (1974).
- [5] BRAAKSMA, B. Critical Dynamics of the Bonhoeffer - van der Pol equation and its chaotic response to periodic stimulation. *Physica. D: Nonlinear Phenomena* 68, 2 (1993), 265 – 280.
- [6] BRAUER, F., AND NOHEL, J. A. *Qualitative Theory of Ordinary Differential Equations: An Introduction*. W. A. Benjamin, Inc., New York, 1969.
- [7] BRAY, W. O. Lecture 6: The Fitzhugh-Nagumo Model. Online Lecture.
- [8] BROWNE, M.W. In Heartbeat, Predictability Is Worse Than Chaos. <http://www.nytimes.com/1989/01/17/science/in-heartbeat-predictability-is-worse-than-chaos.html>, January 17 1989.
- [9] BURDEN, R. L., AND FAIRES, J. D. *Numerical Analysis: Eighth Edition*. Brookes/Cole, Belmont, CA, 2005.
- [10] CAIN, J. W. Taking Math to Heart: Mathematical Challenges in Cardiac Electrophysiology. *Notice of the AMS* 58, 4 (April 2011), 542 – 549.
- [11] CARDIAC LIFE PRODUCTS, INC. NYSAED. <http://www.nysaed.com>, 2011.
- [12] CHEN, J. ET AL. High-frequency periodic sources underlie ventricular fibrillation in the isolated rabbit heart. *Circulation Research: Journal of the American Heart Association*, 86 (2000), 86 – 93.
- [13] FITZHUGH, R. Thresholds and Plateaus in the Hodgkin-Huxley Nerve Equations. *The Journal of General Physiology* 43 (1960), 867 – 896.
- [14] FITZHUGH, R. Impulses and Physiological States in Theoretical Models of Nerve Membrane. *Biophysical Journal* 1 (1961), 445 – 466.
- [15] HEART RHYTHM FOUNDATION. Sudden Cardiac Arrest Key Facts. <http://www.heartrhythmfoundation.org/facts/scd.asp>.
- [16] IZHIKEVICH, E. M. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. The MIT Press, Cambridge, MA, 2010.
- [17] KOSTOVA, T., RAVINDRAN, R., AND SCHONBEK, M. Fitzhugh-Nagumo Revisited: Types of Bifurcations, Periodical Forcing and Stability Regions by a Lyapunov Functional.
- [18] KUZNETSOV, Y. A. *Elements of Applied Bifurcation Theory, Second Edition*, vol. 112. Springer, New York, 1998.
- [19] LOGAN, J. D. *Applied Partial Differential Equations*. Springer Science + Business Media, LLC, New York, 2004.
- [20] LORENZ, E. N. *The Essence of Chaos*. University of Washington Press, Seattle, WA, 1993.
- [21] LYNCH, S. *Dynamical Systems with Applications using Maple, 2nd Ed.* Birkhäuser, Boston, MA, 2010.

- [22] MORRISON, F. *The Art of Modeling Dynamic Systems: Forecasting for Chaos, Randomness and Determinism*. Dover Publications, Inc., Mineola, NY, 2008.
- [23] NAGLE, R. K., SAFF, E. B., AND SNIDER, A. D. *Fundamentals of Differential Equations and Boundary Value Problems, 5th Ed.* Pearson Education Inc., Boston, MA, 2008.
- [24] POOLE, D. *Linear Algebra A Modern Introduction*, thirded. Brookes/Cole, Boston, MA, 2011.
- [25] RANGARAJAN, G. Lyapunov Exponents without Rescaling and Reorthogonalization. *Physical Review Letters* 80 (1998), 3747 – 3750.
- [26] ROCSOREANU, C., A. GEORGESCU AND N. GIURGITEANU. *The Fitzhugh-Nagumo Model*, vol. 10. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2000.
- [27] SEARS, F. W. *University Physics*, 5th ed. Addison-Wesley Publishing Company, Reading, MA, 1977.
- [28] SKELTON, B. Personal Interview. ZOLL Medical Corporation, Mar. 25 2010.
- [29] STROGATZ, S. H. *Nonlinear Dynamics and Chaos*. Perseus Books Publishing, LLC, Cambridge, MA, 1994.
- [30] TAUBES, C. H. *Modeling Differential Equations in Biology*. Prentice Hall, Inc., Upper Saddle River, NJ, 2001.
- [31] THOMPSON, D. W. *On Growth and Form: The Complete Revised Edition*. Dover Publications, Inc., New York, 1992.
- [32] VAN DER POL, B., AND VAN DER MARK, J. The heartbeat considered as a relaxation oscillation, and an electrical model of the heart. *The London, Edinburgh and Dublin philosophical magazine and journal of science* 6 (1928), 763 – 775.
- [33] WINFREE, A. T. *The Geometry of Biological Time*. Springer - Verlag New York, Inc., New York, 1980.
- [34] WOLF, A. Determining Lyapunov Exponents from a Time Series. *Physica* 16D (1985), 285 – 317.

Investigation into the Combined Affects of Tumor Necrosis Factor-Alpha, Tamoxifen and 5-Azacytidine on HTB-26 Breast Cancer Cells

Gregory Roloff and Jessica Bosch

Abstract

The primary goal of modern cancer therapeutic regiments is to develop treatments that are damaging to tumor cells but maintain the health of normal somatic cells. The “cocktails” of drugs used in current cancer treatment can interact in a mechanism that is not adequately understood. However, it is sometimes possible to use combinations of different chemicals to achieve the desired effect of killing the most cancer cells possible while preserving the patient’s other rapidly dividing cells. Our work to this point has been focused on two chemicals, sulforaphane (SFN), a organosulfur compound found in cruciferous vegetables, and Tumor Necrosis Factor- α (TNF), a cytokine involved in regulation of the immune response. The scope of our research is centered around apoptosis, or programmed cell death. In this project, we tested the effects of combining SFN treatment with varying amounts of TNF- α based on prior research from the O’Donnell lab showing the combination of SFN and TNF- α together cause apoptosis despite the fact that TNF- α alone is an insufficient inducing agent. In our preliminary experiments using cytotoxicity assays, we did find increased killing when the drugs were used in combination. Experiments using caspase 3 apoptosis assays showed an increase in apoptosis with the combined treatment due to activation of the caspase 3 pathway. Moving forward, we have begun to investigate how altering gene regulation proteins might sensitize the cells to apoptosis induction via the use of 5-Azacytidine, a DNA methyltransferase inhibitor.

Introduction

The cell line we work with is HTB-26, derived from a 51 year old Caucasian female patient with breast cancer in 1979 (Cailleau). It can be described as an adenocarcinoma of the mammary gland that is derived from epithelial tissue. This cell line is tumorigenic in immunocompromised mice. One of the drugs we worked with is SFN, an isothiocyanate derived from broccoli and other cruciferous vegetables (Figure 1) (Jakubikova). Tumor necrosis factor α is a cytokine involved in

inflammation. Recent research has shown that SFN and TNF- α may have synergistic effects when used together. Previous research has shown that TNF- α may activate NF- κ B, an anti-apoptotic transcription factor (Figure 1).

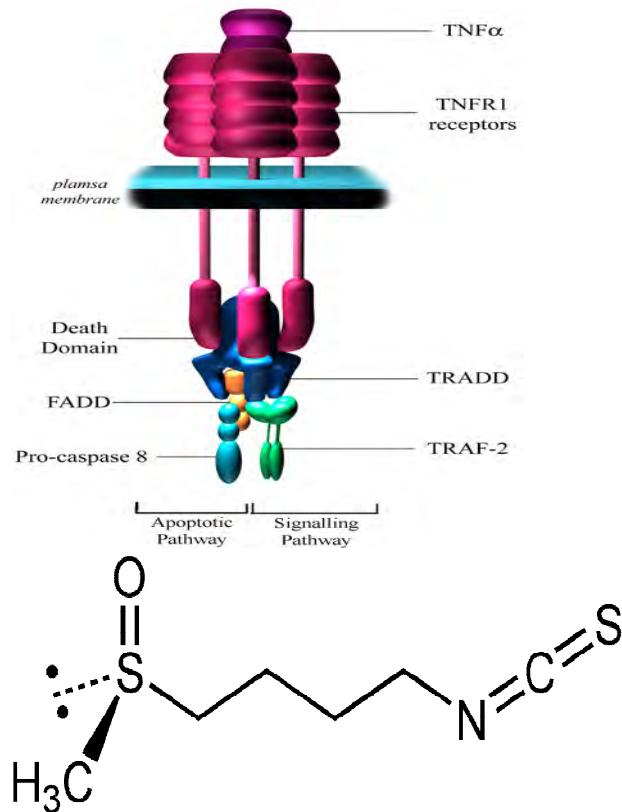


Figure 1. Membrane domain of TNF pathway and chemical structure of sulforaphane

Apoptosis is the process of programmed cell death and cancer cells often have mutations that prevent them from undergoing entering this process. TNF- α activates procaspase 8 which activates the executioner caspases, including caspase 3. Sulforaphane has also been shown to induce apoptosis through activation of reactive oxygen species-dependent caspase-3. This combination of drugs actually has a more lethal effect than either drug alone, leading to more apoptotic cells. Developing hypothesis look to test the levels of cell proliferation after first sensitizing the cells with 5-Azacytidine, a methyltransferase inhibitor (Figure 2.)

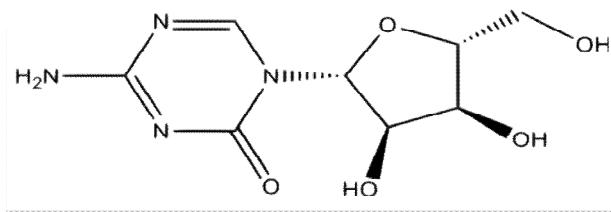


Figure 2. Chemical structure of 5-Azacytidine.

Materials

Dulbecco's PBS (Phosphate Buffered Saline), Trypsin, RPMI (Rochester Park Memorial Institute), L-15 Medium (Leibovitz), L-15, FBS (Fetal Bovine Serum), Glutamax, TNF- α , SFN, Annexin V Kit (10x Binding Buffer, PI, Annexin, Distilled H₂O), Caspase-3 Assay (Lysis Buffer, 2x Reaction Buffer 3, DTT, Caspase-3 calorimetric substrate), CyQuant Assay (CyQuant GR dye, Cell-lysis buffer, DNA standard). MTT, isopropanol, 1N HCl.

Methods

Cyquant Assay : A 96 well plate was set up after cells were cultured and incubated. Each well received 100 μ L of media. SFN was kept constant in all wells through column 10. The bottom four rows received varying amounts of TNF. The plate was incubated at 37° C for 72 hrs and excess fluids were removed by flicking and blotting. A CyQuant assay was prepared, and added to each well and the fluorescence was measured. These experiments allowed us to verify the dosages to use in future experiments.

Caspase 3 Assay: Cells were treated with nothing, SFN, tumor necrosis factor- α or a combination of drugs for 72 hours. Cells are collected by centrifugation at 250G for 10 minutes. The cells were lysed and then centrifuged. In a 96 well plate, 50 μ L of each sample was added to each well along with 50 μ L of 2x Reaction Buffer 3 and 5 μ L of Caspase-3 colorimetric substrate . The plate was incubated for 1-2 hours at 37° C and absorbance was read at 405 nm.

MTT Assay: a 96 well plate was set up to measure cell proliferation. Cells from the breast cancer line HTB-26 were exposed to SFN alone, SFN and

various amounts of TNF, or nothing. The cells were incubated for 72 hours. MTT was prepared and 20 μ L was added to each well and incubated for 5 hours. The supernatant was removed and 100 μ L of a mixture composed of 9.6 mL isopropanol and 0.4 mL HCl (1N) was added before the plate was read for fluorescence at 585 and 690 nm.

Results

Cytotoxicity experiments for HTB 26 cell lines revealed that the concentration of drug that showed the desired amount of killing was 0.035 mM, which ultimately came out to be 0.5mL of 0.28mM of SFN for every 4mLs of cells. The MTT cell proliferation assay detected an additional 7% killing in the presence of TNF, on top of the 57% caused by SFN alone. As the concentration of TNF increased, the amount of killing followed proportionally (figure 3). Results from the apoptosis assay, as depicted in figure 4, suggest that TNF and SFN may have a synergistic effect in inducing apoptosis, since the combination induced higher levels than the additive effects of the drugs alone. The CyQuant Assay illustrates that TNF alone is a poor killing agent, obtaining only a 6% increase from the cell control. However, TNF supplements the strong killing induced by SFN by an additional 6-7% (Figure 5).

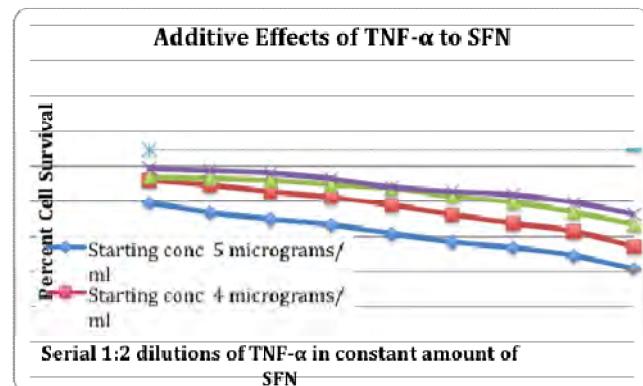


Figure 3. Graph depicting cell survival of HTB-26 cells when incubated in uniform amounts of sulforaphane, while varying the amounts of TNF in the experimental rows. Different doses of TNF are distinguished by separate lines

Discussion

Overall data provides support for the notion that SFN causes cytotoxicity in HTB-26 cells and that a combination of TNF- α and SFN has a greater apoptosis inducing effect. This can have great clinical implications if it is able to kill cancer cells more effectively in vitro. However, treatment would have to be relatively safe to the body's non-cancerous cells, which divide at various rates. The new direction we are following is to determine whether 5-Azacytidine treatment will first weaken cells so they will be more sensitive to apoptosis induction with Sulforaphane and Tumor Necrosis Factor- α . Our further studies will examine the influence these experiments will have on MRC-5 human fibroblasts (non-cancerous cells) and how 5-Azacytidine might enhance previously obtained results.

References

Callieu, R et al. Breast tumor cell lines of metastatic origin: preliminary characterization. *Cancer Res* 39:919-22, 1979.

Jakubikova, Jana. Bao, Yongping. Sedlak, Jan. "Isothiocyanates Induce Cell Cycle Arrest, Apoptosis and Mitochondrial Potential Depolarization in HL-60 and Multidrug-resistant Cell Lines." *Anticancer Research* 25 (2005) : 3375 – 3386.

Acknowledgements

We would like to thank the Geneseo Research Foundation for financing our work, other members of our lab for their assistance and most of all, Dr. Robert O'Donnell for his constant support and encouragement.

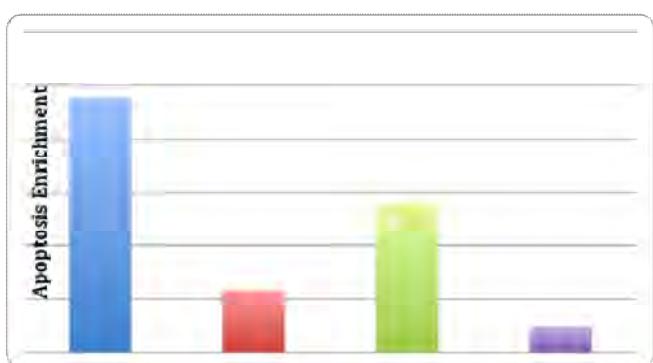


Figure 4. Apoptosis Enrichment of HTB-26 cells based on all experimental treatments of SFN, TNF and a combination of both vs. a cell control that received no drug.

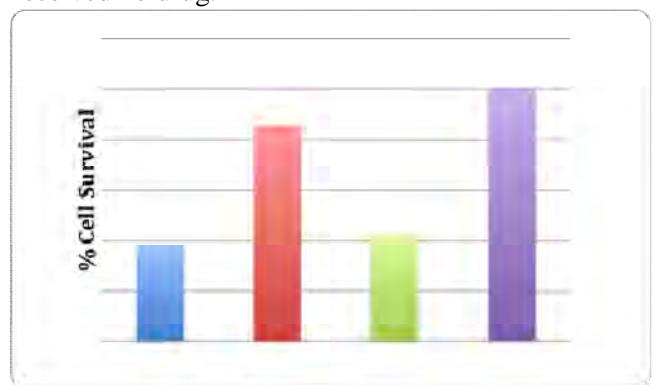


Figure 5. Results from CyQuant Assay for cell survival of each experimental group.

Free Radical - Interactive Game Design

Marcos Davila

When learning the basics of computer programming, there is no finer language to start with than Java. It's a language with a lot of potential which is expanded upon with the inclusion of libraries – dynamically loadable code that can be called at runtime. To a person looking to craft a video game simulation in Java, the most important libraries to be concerned with are the ones that handle graphical user interface and input from the user. With complex interaction between the user and the computer, very simple games can be created in a relatively small period of time. Free Radical, a game simulation programmed entirely by an undergraduate Computer Science major at SUNY Geneseo, would be one of countless such games.

There's a lot more to creating a video game – and programming in general – than what meets the eye. The hardest part is always the starting point. Without an idea and a roadmap, there's no way to achieve the end result. Start with a roadmap and no idea and the resulting program will be little better than alphabet soup – methods thrown together with no idea as to why they're needed or what their overall purpose is. Omit the roadmap and start with an idea and the result will be lines and lines of code that work improperly and are riddled with bugs and errors. There are a few things that need to be worked through before programming could begin. The first question that needed an answer was the simplest: what kind of video game was Free Radical going to be?

I decided on making it a space shooter game, modeling it after *Space Invaders*™ but with a faster pace and a steeper difficulty curve. This immediately called for the creation of a panel class, classes to take care of ships and projectiles and enemies. These classes naturally gave birth to necessary methods and data structures such as movement, storage of enemy ships in the memory of the computer and hit detection between projectiles, ships, the edges of the “playing field” and ships and projectiles. The inclusion of necessary methods gave birth to the inclusion of useful methods – methods that wouldn't have an effect on the overall game but would make the end result look much more cohesive. This led to the inclusion of a scrolling background, a counter to keep track of “points” that the player earns, lives

for the player, power-ups, background music, detailed sprites for the projectiles, ships and power-ups, and so on.

The first question was the most fundamental question, but the second was just as important: how big is Free Radical going to be? Is it going to be one level, two levels, ten levels, or some number in between? This question poses some more obstacles for a programmer to overcome. If there are going to be additional levels, how will it be handled? Will the current panel refresh? Will a new panel pop up in its place and discard the old one? How will the player transfer progress from one level to the next? Plenty of possibilities were available, each more useful in certain situations than others. After eventually settling on programming for two levels, I decided to create an object that upon collision would teleport the player to the next level.

The final major issue of any aspiring game programmer would be that of portability. The goal is to create as much content as possible while keeping the program at a manageable size. The first few demos of Free Radical ranged from 150 to 200 MB in space, which was a large amount of space for such a small game. The final product sits at 30 MB. The large reduction in storage space stems largely from the implementation of the background music. Java comes with the ability to play back WAV files – lossless files that promise high quality audio fidelity in exchange for hogging up a ton of space on your hard drive. The most common format of playback stems from MP3 files, which Java does not natively support. Now I was at a crossroads – do I want background music and a large program, no background music and a small program, or could I have both the background music and a small program size?

This problem highlights the concept of revision. A programmer must be content with knowing that whatever they're writing now will probably need to be rewritten within the next hour. Hardly any code that is ever written at the beginning will remain as is when the program is ready to be published. A programmer must also become fast friends with their search engine of choice, as they will be surfing the Internet time and time again looking for better ways to approach a problem, or looking for a bug fix so they can progress onwards, or looking for an explanation of

what a method does within a program. I eventually found a way to include both the background music at a smaller file size, but the steps to go about it were not intuitive. This highlights one final and very important idea – a programmer must understand how the algorithm works. There are plenty of ways to approach a problem but different solutions are more evident to certain programmers than others. Very often, you'll find that someone else has a better way of approaching a problem and adopting that is fine *as long as you understand what's being done and why*. The only thing worse than debugging lines of code that you wrote is debugging lines of code you don't understand.

The major misconception about programming video games is that it's 90% fun and flashy stuff and 10% boring details that no one wants to get into. It's a fluid, dynamic process that grows upon itself where every step is equally as important as the next and no parts can be skipped and no corners can be cut. The secret to programming is that it's not about the end product – it's about how you get there.

Bibliography

Horstmann, Cay S. and Gary Cornell. *Core Java™ Volume I – Fundamentals Eighth Edition*.
Stoughton: Courier. 2009.

Lorimer, R.J., "MP3: Play MP3's from Java with Javazoom." *Javalobby*.
<http://www.javalobby.org/java/forums/t18465.html> (8 Jul 2011).

Teaching Students of Other Languages

Kelsy Cocozzo

This study was conducted in order to investigate the most efficacious methods for teaching students who speak a language other than English at the secondary level. Though districts are mandated by federal and state governments to take responsibility for the achievement of these students, there is no recommended or mandated program model for teaching language or content to students who are native speakers of a language other than English. A literature review and interviews of a professional development coordinator and three teachers in the Rochester area were conducted to identify the best methods for teaching speakers of other languages and to determine the models that districts are implementing. All names of interviewees have been changed and school names have been omitted in order to protect the privacy of the teachers and school districts in which they are employed. By analyzing research and speaking with current educators, findings suggest that there is often a disconnect between what is accepted as the best practices and what is being implemented in districts. However, the results suggest that models of implementation vary based on whether a school is located in a rural, suburban or urban district.

According to the National Center for Educational Statistics (2010), the number of school-age students that speak a language other than English at home has risen from 9 to 20.5 percent of the total population, or from 3.8 to 10.9 million students. Of these 10.9 million students, 2.7 million, or 5 percent of the 20.5 percent, speak English with difficulty (See Appendix 1). Because these students are not fully proficient in the English language, they are referred to as English Language Learners (ELLs) or Limited English Proficient (LEP) and are enrolled in programs that facilitate the development of their language skills. Because approximately 1 million immigrants move to the United States each year, estimates from 2008 predict that the number of first-generation immigrants will increase to 42 million by 2025 from 25 million in 1996. In addition, the combination of first- and second-generation immigrants will comprise one-third of the population by 2025 (National Center for Education Statistics [NCES], 2008).

The numbers of ELLs are not evenly distributed throughout the United States. According

to the NCES (2010), ELLs make up 8 percent of the student population in the West while only 3 percent of students in the Midwest are ELLs. The states with the highest number of ELLs are not located within the same region. The six states with the largest ELL population are California, Texas, Illinois, New York, Florida, and Arizona, respectively (See Appendix 2). Appendix 2 shows detailed statistics of the populations of school-age ELLs by state and region. The growth of the population of ELLs is also not consistent across the United States. From 1994/95 to 2004/05, the states with the most rapid growth in population of ELLs were Indiana, Kentucky, and South Carolina with an increase of between 400 and 714 percent. Zero of the six previously mentioned states with the highest number of ELLs experienced more than a 100 percent rate of growth (See Appendix 3). The approximate percentage of growth of all other states are shown in Appendix 3. This trend shows that the population of ELLs is beginning to spread throughout the United States rather than being isolated in select states.

Approximately 90 percent of recent immigrants to the United States come from countries where the primary language is one other than English (Echevarría, Vogt, & Short, 2010, p. 5). According to the National Center of Education (2010), of the 2.7 million students labeled as ELLs, 75 percent (1,996,000 students) spoke Spanish, 12 percent (311,000 students) spoke Asian or Pacific Islanders languages, 10 percent (279,000 students) spoke other Indo-European languages, and 3 percent (87,000 students) spoke a language not mentioned (See Appendix 4). These statistics show that ELLs are a very diverse group of students despite their grouping under the same term. Beyond the language spoken, the proficiency levels of the students in their native language and English vary greatly from student to student. It is important for teachers to take this into consideration when planning and implementing instruction.

In order to regulate the education of ELLs, Title III of the No Child Left Behind Act (NCLB) was passed. The act attempts to provide guidelines for schools in implementing programs that increase the proficiency level and content knowledge of students. Due to the current standards-based reform, it holds schools accountable for the

performance of ELLs which includes passing all state examinations. The NCLB Act also set up the State Formula Grant Program to regulate the funding that schools receive from the government for programs ELLs. Each State Education Agency (SEA) that creates a program is approved by the Department of Education receives a formula grant based on the number of ELLs in the state. The SEA then administers subgrants based on the number of ELLs to Local Education Agencies (LEAs) when they have their program approved by the SEA (No Child Left Behind, 2008). Although the act does promote academic achievement of ELLs by mandating that they pass high-stakes test, this is not necessarily beneficial for ELLs. The tests are meant to be tests of their content knowledge but are, in reality, a test of their language skills because they are written for native English speakers. Because they are struggling to develop their skills in reading and writing, they may not be able to accurately understand the task nor be able to adequately express their understanding of concepts (Echevarría, Vogt, & Short, 2010, p. 9). Since NCLB has been passed, the percentage of ELLs who do not graduate from high school has increased to 31% because they cannot pass the high-stakes tests (Echevarría, Vogt, & Short, 2010, p. 5). Although NCLB holds schools accountable for the progress and performance of ELLs, schools view a large population of ELLs as undesirable because it will hurt the accountability statistics of the school (Rance-Roney, 2009, p. 33). NCLB was meant to improve the educational system and increase academic achievement for students. However, it became more difficult for ELLs to obtain a high school education because their language deficit does not allow them to demonstrate their knowledge of content.

NCLB holds schools accountable for the performance of ELLs but does not require any professional development of teachers that work with ELLs. It is the responsibility of schools to provide professional development to teachers or for teachers to seek out professional development independently. Due to the lack of funding for programs and time constraints, the majority of teachers have not obtained sufficient, if any, training for working with ELLs. In 2000, 41.2 percent of teachers had taught ELLs but only 12.5 percent had eight or more hours of professional development within three years (Echevarría, Vogt, & Short, 2010, p. 4). ELLs cannot succeed in an environment where teachers are unaware of how to differentiate instruction to make content accessible

to them. Continued professional development is essential for teachers of ELLs because it will give students the tools that they need achieve the high standards set by NCLB.

In addition to NCLB, New York State has specific requirements and expectations for the education of ELLs stated in Part 154 of the Commissioner's Regulations. New York State alone has 881,000 students who speak a language other than English at home and 203,000 students who are considered to be ELLs (See Appendix 2). Under Part 154, students who report speaking a language other than English take the Language Assessment Battery-Revised (LAB-R) to determine whether they will be identified as an English Language Learner at the beginning of the first year of formal schooling. Each subsequent year, ELLs are given the New York State English as a Second Language Achievement Test (NYSESLAT) to determine the proficiency level of the students and whether they test out of free-standing English as Second Language services (Commissioner's Regulations, Part 154, 2007, p.2). The NYSESLAT determines how many units of ESL and ELA the students will receive with each unit being 180 minutes per week. Because this study concentrates on grades 7 through 12, only units for these grades will be considered. For an ESL program in grades 7 and 8, beginning and intermediate proficiency levels receive two units of ESL and no units of ELA. For an ESL program in grades 9 through 12, beginning level proficiency students receive three units of ESL and no ELA, intermediate level students receive two units of ESL and no ELA, and advanced level students receive one unit of ESL and one unit of ELA. For a bilingual program, students must take one unit of Native Language Arts (NLA) in addition to the same number of ESL and ELA credits as in an ESL program (See Appendix 5). All ESL classes must adhere to the NYS Learning Standards for ESL, which include five standards, various performance indicators under each standard, and sample classroom tasks by proficiency level (New York State Education Department, 2004).

Though Part 154 gives clear methods of language assessment and directly states the amount of instruction each student should receive, the document does not give much more specific information concerning the education of ELLs. All instruction must be based on the ELA and ESL standards and integrate content (Commissioner's Regulations, Part 154, 2007, p. 3). This, however, is a requirement for the free-standing ESL program

and not necessarily need to be considered by the content teachers with ELLs in their classes. Part 154 recognizes only free-standing ESL programs and bilingual programs. If a school district has 20 or more students who speak the same native language and are in the same grade level, the school must implement a bilingual program. If the school does not have these numbers, a free-standing ESL program or a bilingual program may be used (Commissioner's Regulations, Part 154, 2007, p. 9). The document does not give any details about how to develop, implement, or maintain either type of program. However, Part 154 does hold schools accountable for the performance of ELLs by making districts report information about the program implemented for ELLs and their performance. Districts must report to the state how they evaluate ELLs on a yearly basis, the philosophy of education for teaching ELLs, the services offered to ELLs, a description of how the program the district has chosen was constructed and is managed, and the results of the yearly evaluation of ELLs (Commissioner's Regulations, Part 154, 2007, p. 7). Though holding districts accountable for the performance of ELLs is important for tracking progress of ELLs, it would be more beneficial if districts were given concrete and research-based information about how to instruct ELLs because it would give strategies that are proven to work with ELLs rather than forcing districts to build their own programs. The Performance Indicators and sample classroom tasks contained within the NYS Learning Standards for ESL offer the most information for creating an operative program for ELLs.

As shown by the analysis of federal and state government documents, there is a lack of readily available resources and information that is government approved about creating a program for teaching ELLs. Therefore, it is the responsibility of districts to decide what they believe will be the most effective program and develop it independently. According to Joe Smith, a professional development coordinator at Bilingual ESL Technical Assistance Center (BETAC) that provides support for five Boards of Cooperative Educational Services (BOCES) in the city school district including Monroe 1 and 2, Genesee Valley, Dwayne Finger Lakes, and the Greater Southern Tier which encompasses 88 districts, three-quarters of these districts have some type of ELL population. BETAC works to develop program models and provide professional development on a school by school basis by assessing what the school

is doing and determines what they need to create a more successful program. However, BETAC only gives advice and support to schools if the school directly requests assistance in order to avoid overbearing interference (J. Smith, personal communication, February 18, 2011). Although BETAC and BOCES have the resources to assist districts in implementing programs for ELLs, it is still the responsibility of the districts to be proactive and reach out to and take advantage of these services.

One way that districts begin to develop program models is to determine the needs of ELLs and then structure a program that will meet these needs. However, this is not as simple as it seems because ELLs are not a homogeneous group of students although they are grouped together. They have different proficiency levels and background knowledge due to interrupted schooling, home language, native country, immigration status, cultural understanding, socioeconomic status, and time spent in the United States. ELLs must overcome factors such as lack of literacy in any language, sporadic education, little to no knowledge of English, as well as other elements of their identity as an ELL (Rance-Roney, 2009, p. 34). For example, it is evident that ELL status is related to poverty because 10 percent of students at or below the poverty line and 8 percent of students near the poverty line are considered to be Limited English Proficient whereas only 3 percent of the population that is considered non-poor, or having 200 percent or more of the poverty threshold, are labeled as LEP (NCES, 2010). Because of these factors, meeting the needs of ELLs is not always a simple or easy task for districts. In general, ELLs benefit from school reforms that would be beneficial for all students such as reorganization and improvement of the curriculum and increased professional development for all teachers (Rance-Roney, 2009, p. 34). Though these reforms would cost the districts a significant amount of time and funding, they would benefit ELLs as well as the entire district.

The task of meeting the needs of ELLs should not be left to the ESL teacher alone. According to Emily Davis, the role of an ESL teacher is to provide the specific ESL instructional periods to ELLs and to act as a resource for content teachers who have ELLs in their classes. ESL teachers are professionals with extensive knowledge about language acquisition and research-based methodology for teaching ELLs (E. Davis, personal communication, March 24, 2011). All of the

professionals who were interviewed agree that ESL teachers should never act as a tutor for content area courses or as an aid in the room of content area teachers (J. Smith, personal communication, February 18, 2011; M. Adams, personal communication, March 4, 2011; J. Brown, personal communication, March 9, 2011; E. Davis, personal communication, March 24, 2011). ESL classes should not be structured as a study hall where ELLs come to receive assistance on assignments for their core classes. Instead, ESL teachers could help students with their content courses by recognizing the skills that they are trying to build and planning instruction to develop these skills and academic processes. For example, if students are writing an essay in their Global History class, the ESL teacher should teach lessons and plan activities that guide students through the processes of doing research, making graphic organizers, writing outlines, constructing thesis statements, and other aspects of essay writing. ESL teachers often fall into the trap of acting as an aid in a mainstream content class when push-in situations are implemented. For this reason, most ESL teachers at the secondary level avoid push-in situations because they are generally not an effective use of time (J. Smith, personal communication, February 18, 2011). If ESL teachers do decide to push into a classroom, the purpose must be made clear, concrete roles must be established, and the lesson must be co-planned in advance so that it is more of a co-teaching situation than an ESL push-in format. If ESL teachers are not used for their proper purpose, it is a waste of a valuable resource for both ELLs and other teachers.

The role of an ESL teacher can differ based on grade level, school district, program model chosen, and the goals of the students. For example, an adult education ESL course will look significantly different than the ESL course in a high school because the ages and desired outcomes for the ELLs are different (E. Davis, personal communication, March 24, 2011). However, the core responsibilities of ESL teachers should remain the same. ESL teachers are first and foremost language teachers with some aspects of an ELA teacher because their main focus should be teaching the English language and literacy during ESL periods (M. Adams, personal communication, March 4, 2011). Though the language is taught through relevant content, language is the focus during ESL periods. ESL teachers should focus mainly on language development in the four essential skills of listening, speaking, reading, and writing and to create content-based language

instruction that is appropriate for the proficiency and grade level of the ELLs (J. Smith, personal communication, February 18th, 2011). In general, teachers should teach ELLs in their areas of specialization, meaning that ESL teachers should concentrate on teaching language and literacy and content area teachers should focus on teaching content (J. Brown, personal communication, March 9, 2011). However, these two roles are not completely independent of one another because there should always be content present in language development classes and content teachers are responsible for emphasizing content-related vocabulary and making content linguistically comprehensible for ELLs.

Districts, schools, teachers, and students must be clear on the difference between ESL classes and content classes and the role of both types of classes. ESL courses are considered to be English Language Development (ELD) courses, which should focus primarily on language but also have a content-based element (Goldenberg & Coleman, 2009, p. 11). Though ESL classes are structured based on the NYS Standards for ESL, content should still be present in the lesson in order to contextualize the language but these classes do not necessarily have to focus on the same exact content as the mainstream content courses but should relate to some type of content-related information that would be useful to the students in mainstream classrooms. For example, if a 10th grade Biology class were studying photosynthesis and other plant processes, the ESL class could study a general scientific topic such as the scientific method. Therefore, if there were students in different grades in the class, the information would still be applicable to their coursework. ESL classes, however, are not and should not be confused with standards-based content instruction (Goldberg & Coleman, 2010a, p. 64). Though language is not the focus of content classes, content teachers do need to think about the language needs of ELLs. Content teachers should collaborate with ESL professionals in order to adapt materials, assignments, instruction and any other language oriented elements of the class. The academic and language roles and expectations of ESL and content classes should be clearly defined so that all educators of ELLs can do their jobs effectively and thoroughly.

ESL professionals, content teachers, counselors, and administrators should form teams that collaborate to create instruction for ELLs. The schedules should be made so that all members of the team have planning periods together on a

regular basis (Rance-Roney, 2009, p. 34). This would allow for the planning of thematic units and discussion of the individual needs of students. Cross departmental collaboration and planning is essential even if it is as minimal as having one planning period for the ESL teacher and content teachers or as much as breaking down departmental lines in order to fully integrate content and language (Spaulding, Carolino, & Amen, 2004, p. 12). Although ESL teachers are a resource that content area teachers should use to learn more about educating ELLs, oftentimes ESL teachers must take the initiative to reach out to content area teachers and suggest collaboration (J. Brown, personal communication, March 3, 2011). ELLs benefit from thematic units across various subject areas because it gives meaningful context to academic language and content and allows them to practice with specific concepts and words throughout the day rather than for only one period per day. In addition to cross-departmental collaboration, complete records should be maintained for each student so that all teachers have access to information that could affect the academic success of students. These records should contain updates on information such as yearly progress, proficiency level, language experience, and family factors (Rance-Roney, 2009, p. 36). In order to provide appropriate support and instruction to ELLs, communication about their progress and current needs should be open and frequent.

In terms of language needs, ESL programs must focus primarily on building the academic literacy of ELLs. In fact, the primary goal of all ESL programs should be the development of literacy because it is a key component in all content area classes. If ELLs do not have a high level of academic literacy, they will not be successful in mainstream content courses. ELLs should be taught literacy skills explicitly so that they can eventually independently apply the techniques to other contexts and subject areas (Coleman & Goldenberg, 2010b, p. 108). Literacy skills do not help ELLs if they can only use them with the direction of a teacher or in the specific situations that they were introduced. At the beginning and intermediate levels, the English language proficiencies of ELLs may be a barrier that prevents them from being able to comprehend, use, and internalize these strategies. Therefore, it would be beneficial to teach literacy strategies to students in their native language if the resources are available. It is necessary to assess the literacy skills and strategies that ELLs know and implement in

their native language in order to determine what they still need to be taught. By using the native language of the students, teachers can assess their true knowledge of language and development of literacy skills without the added factor of considering the role of limited English proficiency status. In general, literacy professionals agree that students who are taught literacy strategies in their native language are more likely to make progress with academic tasks in English than students who are taught these strategies in English (Mora, 2001, p. 151). However, teaching literacy strategies to ELLs in their native language is not sufficient because they often cannot take skills learned in their native language and use it in English without proper guidance. Teachers should constantly monitor and assess the transfer of literacy skills from the students' native language to English in order to determine which skills have been transferred, which still need to be transferred, and which skills are completely lacking and must be developed in both languages (Mora, 2001, p. 156). ESL and content area teachers must be aware of the process that ELLs are going through and guide them along the way until they show that they can use the literacy strategies in English confidently and skillfully.

Unfortunately, there are certain obstacles that stand in the way of teachers attempting to strengthen the literacy skills of ELLs. Though research conclusively proves that skills should be taught in the native language of the student, this is not always possible. Some schools do not have any bilingual teachers and, therefore, do not have the proper resources to provide instruction in the native language of ELLs. If the ELL population of a school speaks multiple languages, it is extremely difficult to provide native language instruction to all students due to a lack of trained professionals, resources, and time for planning. Having various languages within one class can create an unequal learning environment for ELLs if all instruction is not carefully planned due to a disparity of support based on native language (Coleman & Goldenberg, 2010, 2010b, p. 110). For example, if an ESL class was comprised of ELLs who spoke Spanish, Arabic, French, and Japanese and the school had a bilingual teacher that spoke Spanish, it would be unfair to the other students if only the Spanish-speaking students were provided with native language instruction in addition to the supports provided to the rest of the class and the other students were not given any extra support. The support given does not necessarily have to be equal

but it must be equitable so that all ELLs have the same opportunity to learn and improve. Literacy skills are the most important aspect of the education of ELLs because a high level of literacy is an essential life skill that will allow ELLs to graduate high school, receive a higher education, and obtain desirable career goals.

In addition to literacy, ELLs have specific needs that correspond to their proficiency levels. Both Basic Interpersonal Communication Skills (BICS) and Cognitive Academic Language Proficiency (CALP) are important for ELLs. BICS represents the conversational language that students would use with peers or adults in non-academic or casual situations. If BICS is not developed in ELLs, teachers should use a combination of explicit instruction and communication through authentic situations to build BICS. Explicit instruction is not sufficient because it does not give students a chance to use and practice what they have learned, an idea that is widely accepted by researchers and leading theorists in language acquisition. The majority of researchers in the field agree that authentic communication is also not enough to move ELLs to advanced and native-like proficiency levels. Using only authentic communication between ELLs and their peers is not beneficial because students may develop a interlanguage pidgin of English that can have many significant transgressions from Standard English. Then these students will fossilize these mistakes, an idea and phrase coined by Larry Selinker who is a major contributor to theories on the acquisition of a second language. Fossilization refers to the learning of incorrect forms during the language acquisition phase. Practicing incorrect or imperfect language with peers will encourage fossilization for ELLs, which is why authentic communication must be supplemented with explicit instruction (J. Smith, personal communication, February 18, 2011). Opportunities for oral communication should be structured so that the students are accomplishing a specific task rather than speaking about random topics without a purpose (Coleman & Goldenberg, 2009, p. 13).

CALP, on the other hand, refers to written and oral academic language that is essential for the success of ELLs in the scholarly realm. Teachers must ensure that ELLs can understand and use CALP so that they can comprehend decontextualized language in reference to content and express their knowledge of the content that is taught (Coleman & Goldenberg, 2009, p. 15). ELLs generally do not perform well on standardized tests because they assess the students' effective use of

CALP, which is where ELLs tend to struggle and fall behind grade-level norms. In practice, it takes ELLs multiple years to develop enough CALP to be able to participate fully in mainstream content courses (Goldenberg & Colman, 2010, p. 64). ESL teachers should emphasize CALP in their classrooms by intertwining language and content. Therefore, ESL classes should teach language through content rather than structuring the class as a modified ELA course (Rance-Roney, 2009, p. 35). Though BICS and CALP are two different categories of language, they are both important for ELLs to develop because they are linked. The progression from BICS to CALP can be seen in Appendix 6, which demonstrates the changes made in the language when moving from conversational language to academic language. Teachers can show ELLs how to take skills and strategies that they use in conversational language and transfer it to CALP (Goldenberg & Coleman, 2010a, p. 62). By asking students to consciously think about the ways of approaching and transferring language skills, teachers are allowing students to actively engage with the inner workings of the language rather than just learning vocabulary and constructions.

Educators must take into account both BICS and CALP when determining the proficiency level of students. There is no concrete research that gives conclusive results about the effects of grouping students based on proficiency level although literature about the education of ELLs does suggest that it may be helpful because it is proven to be helpful to group students by math and reading levels (Goldenberg & Coleman, 2009, p. 14). Because ELLs can be at many different points on the spectrum of proficiency, it is difficult to come up with designated categories. Part 154 uses the results of the NYSESLAT to categorize students as either beginning, intermediate, or advanced levels of proficiency (Part 154, 2007, p. 3). These groups are very broad and within each group there could be a wide distribution of proficiency levels within beginning, intermediate, and advanced. For example, a beginning level student who is just entering the program will have a different proficiency level and needs than a beginner who is about to move to the intermediate level although both ELLs are categorized as beginners (J. Brown, personal communication, March 9, 2011). Grouping students is beneficial within ESL classes because the focus is on language and the teacher will be able to plan differentiation that will facilitate language development of the proficiency levels of each group of students. There are many

challenges of reaching students of different proficiency levels within the same class such as finding applicable content when ELLs may be in various grades and creating time for planning differentiation. ESL teachers must find a common link between the content in various grade levels in order to contextualize the language and make the information meaningful for all ELLs (J. Smith, personal communication, February 18 2011).

Content teachers must be careful with grouping in content area classrooms that contain ELLs. Teachers often group all ELLs together, which does not give them the opportunity to interact with their native-speaking peers. This segregation does not allow ELLs to gain experience listening to and speaking with native English speakers, which could help them improve upon their own English language skills. Another common mistake of content teachers with ELLs in their classrooms is to group ELLs with struggling English only students. This is not effective for the ELLs or the academically struggling students because they are suffering from different issues with learning and must be provided with the correct supports to overcome these obstacles (Goldberg & Coleman, 2009, p. 14). Students should not be grouped together in an academic setting simply because they share the same title, such as ELL, or because they are struggling in school. Students should be looked at as individuals with unique characteristics that should be taken into consideration when grouping them together and differentiating instruction.

Proficiency levels should also be carefully considered when making class schedules for ELLs. As previously mentioned, the number of units of ESL and ELA that each ELL must take is determined by their English proficiency level (See Appendix 5). ELLs also need to be enrolled in content courses, which will most likely be entirely in English and structured for native speakers of English unless ELLs attend a school with a bilingual program. ELLs should not be deprived of the opportunity to take high-level content courses because of their level of English proficiency. Schools need to take into consideration the proficiency level of the students as well as their content knowledge and academic potential. If an ELL has the academic potential and adequate background knowledge to be a member of an advanced course, the school should provide the necessary linguistic supports to the student so that they can be successful in the class. For example, if a 10th grade ELL has had sufficient education in

science and could succeed in a mainstream Biology course, which average English speaking students would take in 10th grade, the students should be able to enroll in that course. The ESL teacher and Biology teacher should collaborate in order to make content comprehensible and provide linguistic assistance to the student. A student should never be placed in a lower-level or remedial class due to their developing language proficiency (Rance-Roney, 2009, p. 37). Many schools offer the option of individualized pathways for students that would accommodate students through extra night classes, summer school, or block scheduling. These options would allow students to spend more time working on their English language development as well as the courses that they will need to graduate. When these options are carried out correctly and for the purpose of giving ELLs more educational time to engage with language and content, individualized pathways are beneficial for ELLs. However, many schools misinterpret and misuse the idea by enrolling ELLs in study halls, life skills courses, or remedial classes in order to decrease their course load (Rance-Roney, 2009, p. 36). This is counterproductive for ELLs because they are already linguistically disadvantaged and the school is making them academically disadvantaged by assigning them to filler classes. These classes will not help ELLs increase their level of English proficiency nor do they fulfill any graduation requirements.

ELLs have multiple linguistic and academic needs that must be met in order to ensure their success in content area classes. Schools must recognize these needs, choose a model that they believe will meet the needs of the ELL population, and implement the chosen model. When a model of instruction is chosen, it should be implemented by all teachers in all classrooms, grade levels, and schools within the district. ELLs benefit most from instruction that is consistent and it helps them assimilate into mainstream classes (NCES, 2008, p. 30). All teachers that work with ELLs should be provided with adequate and continued professional development because most teachers who have ELLs in their classrooms have no training at all (Spaulding, Carolino, & Amen, 2004, p. 6). According to the NCES, teachers that work in districts with greater numbers of ELLs are more likely to have training than teachers that work in districts with few ELLs. This is not acceptable because the needs of ELLs must be met regardless of whether the percentage of the population is high or low. Unfortunately, many small school districts

tend to neglect the needs of ELLs because they cannot justify spending money on programs for a small number of students. Of teachers that have classes with the majority of students being ELLs, slightly less than 75 percent have had some form of training for working with ELLs (NCES, 1997). Both federal and state laws require that teachers provide ELLs with a quality education. However, if these teachers are not required to have any training in working with ELLs, they are not providing ELLs with the best education possible because they cannot accommodate instruction for them. Minimal training such as one day workshops or weekend professional development conferences are not enough time for teachers to be trained in a certain model and then implement it effectively. All four ESL professionals that were interviewed agree that content teachers do not have enough, if any, training for teaching ELLs and, therefore, do not understand how to make content comprehensible for ELLs (J. Smith, personal communication, February 18, 2011; M. Adams, personal communication, March 4, 2011; J. Brown, personal communication, March 9, 2011; E. Davis, personal communication, March 24, 2011).

Because the number of ELLs has been increasing so quickly and immigrants are moving to areas that previously had no ELL population, schools are being forced to set up programs on short notice with few resources and teachers trained to work with ELLs (Echevarría, Vogt, & Short, 2010, p. 5). Therefore, students are not receiving the best quality education because schools are ill-equipped to provide services for the ELL population. Administrators request guidance in how to begin to implement a program and how to appropriate funds and resources for the program. A great amount of research has been conducted on effective instructional practices, which are usually just individual techniques that could be used in the classroom to achieve one specific goal. There are not many guides that combine these individual techniques in order to form a complete and cohesive program model, which could be followed and implemented in a straightforward manner by schools and whole districts (NCES, 2008, p. 11). With the completion of a literature review and interview of ESL teachers and professional development coordinators in the Rochester area, it is evident that a lack of concrete program model is a major obstacle in the implementation of an effective program for ELLs.

After careful analysis of the data collected, there is a clear indication of the most efficacious

methods for teaching ELLs. There is not one best method for all schools because the size and diversity of the population of ELLs must be taken into account, as it is in Part 154. As previously explained, in New York, if a school district has 20 ELLs that have the same native language and are in the same grade, the district must implement a bilingual program (Commissioner's Regulations, Part 154, 2007, p. 9). The majority of researchers agree that additive bilingual education is the best method for teaching ELLs if there is a sufficient number of ELLs that speak the same language. In an urban school district in the Rochester area, a bilingual approach to the education of ELLs is used because 67 percent, or approximately 380 students, in the school are now or have previously been labeled as ELLs and 325 students are still enrolled in the bilingual program. Because 99% of the ELL population speaks Spanish as their native language, the school is able to implement a Spanish-English bilingual program. Two students from Nepal, three students from Vietnam, and one student from Haiti attend the school but are not enrolled in the bilingual program because their native language is not Spanish. These students are enrolled in ESL courses and are given other supports and modifications in place of the bilingual education such as the use of bilingual dictionaries during testing, modified texts, completed graphic organizers, translations of materials, and so on. Beginning level students receive all of their content area courses in Spanish; intermediate level students receive some content courses in Spanish and some in English; advanced students receive all content area courses in English (J. Brown, personal communication, March 9, 2011). Though advanced students receive the majority of their instruction in Spanish, respect and continued development of their native language is maintained by providing students with a class in Native Language Arts.

Additive bilingualism is the process of developing and valuing both the native language and the second language throughout the education of ELLs as opposed to subtractive bilingualism which replaces the native language with the second language and neglects the native culture of the students (Haley & Austin, 2004, p. 62). Additive bilingualism is true bilingualism because it works to develop both languages as part of the students' identity rather than forcing students to leave behind their native language and cultural background as in subtractive bilingualism. The active use and advancement of both languages stimulates cognitive and academic growth (Spaulding,

Carolina, & Amen, 2004, p. 27). Because bilingual programs aim to develop both native and second languages, students should be assessed in both languages (Escamilla and Coady, 2001, p. 43). For example, if an English teacher only looks at an essay from an ELL, the teacher may conclude that the student does not know the information. If the teacher also had the student write an essay in Spanish about the same topic, the teacher would be able to see whether the student was struggling with content knowledge or whether the student's English language skills were preventing the student from expressing his or her understanding of the topic.

Because literacy is the main goal of programs for ELLs, it is important to consider how programs promote literacy and the use of literacy skills. In bilingual programs, literacy instruction is generally given in the native language so that students can apply these skills to English, which can improve the reading and writing of students in their native language and in English (Tinajero & Hurley, 2001, p. 32). In order to close the gap between the literacy of ELLs and the literacy of native speakers of English, bilingual programs should be implemented consistently for five or six years (Rothenberg & Fisher, 2007, p. 18). Evaluating both languages is essential to understanding the student's level of literacy in both languages. Without assessing and comparing the work done by students in both languages, the progression of the biliteracy of students cannot be determined. If both languages are evaluated, teachers can better understand the students' strengths and weakness, linguistic development, and instructional needs, which will allow teachers to plan instruction to close any gaps in linguistic or content knowledge (Escamila and Fisher, 2007, p. 18). In bilingual programs, teachers are able to know all facets of the ELLs' linguistic background, growth, and achievement because the students' linguistic abilities in both languages can be taken into consideration. Bilingual education provides teachers with the most in depth and comprehensive view of the linguistic capabilities of ELLs. In turn, it will encourage bilingual students to obtain academic excellence across the curriculum, complete bilingualism, and cross-cultural intelligence (Spaulding, Carolina, & Amen, 2004, p. 27). If schools choose to institute a bilingual program, it is imperative that both languages are developed and evaluated to ensure that the languages work together to strengthen one another.

There are many trends that have been researched and recorded that suggest that bilingual

programs help ELLs improve their language skills immensely and give them the tools that they need to graduate high school. Researchers have found that the most accurate predictor of academic achievement of ELLs is the number of years of education that they have had in their native language. ELLs who have had more years of native language instruction are more likely to achieve high levels of proficiency at a quicker rate than those who have had few years of native language education (Rothenberg & Fisher, 2007, p. 18). Bilingual education allows students to develop their English language skills while continuing to receive education in their native language so the number of years of native language instruction that they have received keeps increasing and acting as supplemental resources to the acquisition of English. Because the knowledge of students, including ELLs, is assessed by state-mandated standardized tests, programs need to prepare ELLs for these tests, linguistically and academically. Wayne Thomas and Virginia Collier conducted a long-term study from 1985 to 2001 on the achievement of ELLs in relation to the type of program that they were involved in. The study proved that bilingual programs are the only programs that allowed ELLs to reach the 50th percentile in both their native language and English across the content areas. By the time ELLs who participated in bilingual programs graduate high school, students were able to maintain or reach higher levels of achievement (Rothenberg & Fisher, 2007, p. 18). Bilingual programs use the native language of students to ensure the success of students in all subject areas.

Although bilingual programs have many benefits, they do have certain drawbacks. Bilingual programs at the secondary level are uncommon due to the lack of qualified bilingual teachers and problems with scheduling planning periods and classes (Spaulding, Carolina, & Amen, 2004, p. 27). As previously mentioned, in 2000, 41.2 percent of teachers had taught ELLs while only 12.5 percent had eight or more hours of professional development within the last three years, which shows the lack of teachers who are qualified to work with ELLs (Echevarría, Vogt, & Short, 2010, p. 4). At the secondary level, the most qualified teachers would be content teachers who can fluently speak the native language of the students as well as English. However, recruiting bilingual content teachers for every subject in each grade level is troublesome so a consistent bilingual program is difficult to implement and maintain. A

lack of planning periods is a problem because implementing a program where multiple professionals need to meet on a regular basis in order to collaborate and coordinate instruction would take out of school planning, which many teachers are unwilling to do. Another drawback to bilingual programs is that they do not always provoke immediate improvement of the English skills of ELLs. In elementary school, ELLs who choose to leave language support programs initially outperform ELLs in bilingual programs when tested in English. At the middle school level, students in bilingual programs catch up to ELLs who are in mainstream classes. By high school, the achievement levels of students in bilingual classes had exceeded the levels of students from English-only classes (Rothenberg & Fisher, 2007, p. 18). Many schools districts do not take this into account and assume that because ELLs in mainstream classes perform at a higher level than ELLs in bilingual programs, this trend will continue. On a national level, many right-wing government officials, school administrators, and citizens did not want bilingual education because they saw it as unsupportive of America and American ideals. They wanted a program model that would replace the native language of students with English rather than nurturing and developing both languages. Schools petitioned to have bilingual programs transformed into ESL programs in order to force students to assimilate and conform to American culture and the English language and neglect their home language and culture (Haley & Austin, 2004, p. 62). Given these obstacles, it is understandable that not many school districts have complete and consistently implemented bilingual programs from elementary school through high school.

A bilingual program can only be implemented under very specific circumstances and can only be maintained with much funding, planning, and commitment to a school-wide bilingual model. Therefore, for schools that have the resources and a large enough population of ELLs in the same grade that speak the same language, bilingual education would be the ideal choice. However, due to the recent trend of immigrant families settling in communities that were previously homogeneous in language and culture, the majority of school districts do not have the adequate resources or population of ELLs, which means that bilingual education is generally not the most practical or effectual model of teaching ELLs (NCES, 2008, p. 1). The results of a literature review and interview of professionals who work with ELLs suggests that

sheltered instruction is most often the most effective model for teaching ELLs.

Sheltered instruction (SI), also known as Specially Designed Academic Instruction in English (SDAIE), offers courses that meet grade level objectives but teachers modify instruction and language for ELLs. Language development and sheltering techniques should be incorporated into content instruction in order to help ELLs learn content and language simultaneously (Goldenberg & Coleman, 2010b, p. 62). SI meets the language and content needs of ELLs and can be used in a variety of different situations, which makes it easy to implement and adapt to the needs of the students. Sheltered instruction has roots in the findings and theories of researchers such as Stephen Krashen and Tracy Terrell, who are considered to support communicative approaches. Krashen and Terrell supported the idea of language acquisition rather than language learning. Language learning is a very linear and conscious process of studying words, rules, and sentence structure and paying close attention to form whereas language acquisition is a process that involves actively engaging in meaningful interaction using the language to gain greater linguistic abilities (Haley & Austin, 2004, p. 11). During classes teachers should provide as much comprehensible input as possible, which means that students should be able to understand the majority of the teacher's speech but not necessarily all. The input should be slightly above the proficiency level of the students in order to push them to understand more challenging language, which will allow them to improve linguistically in terms of vocabulary and complexity of sentences. Krashen and Terrell believed that the meaning of communication is more important than the form at the acquisition level so educators should plan authentic and creative situations for students to communicatively engage in to practice the language. They also believed that error correction should be kept to a minimum because research does not show that error correction enhances the learning of the language, which is a theory that has since been debated (Haley & Austin, 2004, p. 51). Though theories and practices have evolved from the ideas of Krashen and Terrell, sheltered instruction draws from the basic notion that authentic communication and interaction are essential for ELLs in their acquisition of the English language.

Content-based sheltered instruction is favored because it is aligned with national and state standards. Four of the five New York State

Standards of ESL focus on the use of academic language to communicate, which demonstrates that language and content should be inextricably linked and are equally essential to the success of ELLs (NYSED, 2004). Because SI classes focus on content knowledge, they should be taught by content-area teachers who are trained to work with ELLs and taught how to implement sheltered instruction. The instructional methods used in sheltered instruction are based on those used in both second language and mainstream classrooms, such as the use of visual aids, demonstrations, adapted texts, targeted vocabulary development, and supplementary materials (Echevarría, Vogt, & Short, 2010, p. 15). In general, SI can be considered a compilation of the best teaching practices because it makes use of supplemental materials and other supports to make content more accessible. In fact, content area teachers may be sheltering materials and instruction without consciously thinking about conforming to a specific model simply because they are using good teaching techniques (M. Adams, personal communication, March 4, 2011).

Sheltered instruction provides the best balance between language and content instruction. ESL classes focus on English while using content to practice and use the language. If students are not provided with SI in their content classes, they will not be able to obtain the proper amount of grade level content because they will not be able to decipher the language. Many school districts try to remedy the language barrier by incorrectly placing students into remedial courses where they do not belong academically (Spaulding, Carolino, & Amen, 2004, p.13). SI can act as the bridge from bilingual or content-based ESL to mainstream courses. Therefore, the amount of SI students receive should increase as the students transition out of these programs and into the mainstream classroom (Echevarría, Vogt, & Short, 2010, p. 16). ESL courses should still be provided in addition to sheltered content courses as a supplemental language assistance and development. SI within mainstream classes with native English speakers is beneficial for ELLs because they gain more exposure to English and authentic interaction with native speakers. If the students are in need of extra support, a bilingual aid or co-teacher could provide guidance or assistance to ELLs in their native language (Rance-Roney, 2009, p. 36). This native language support should be used strategically and sparingly so that students do not become dependent on or simply wait for instruction in their native language.

The idea of sheltered instruction is very broad because it encompasses many techniques that modify instruction for ELLs. Therefore, in many classes that implement SI and ESL courses in general, teachers were encouraged to pick activities that they thought would work or found interesting because there were no concrete models to follow (Echevarría, Vogt, & Short, 2010, p. 15). This element of guesswork often results in wasted time and a lack of positive progress for ELLs. Though groups of ELLs have different needs and are composed of diverse learners from various linguistic and cultural backgrounds, certain research-based strategies are proven to be effectual for teaching language and content to ELLs. For this reason, Jana Echevarría, Ph.D., MaryEllen Short, Ed.D., & Deborah J. Short, Ph.D. created the Sheltered Instruction Observation Protocol (SIOP) Model, a model of lesson planning and implementation for sheltered instruction, after a 7 year research study conducted for the Center for Research on Education, Diversity and Excellence (CREDE) from 1996 to 2003 (Honigsfeld & Cohan, 2006, p. 4). The SIOP model was developed because most SI classrooms did not look the same in terms of instruction, activities, and interaction between teachers and students. Therefore, the creators of the SIOP model developed a reliable, effective, and consistent model of SI. According to the creators of the SIOP model, researchers in the educational field, and current teachers, it is the most effective model because it is a form of sheltered instruction that is explicitly explained so that teachers can plan instruction that is directly aligned with the model. Therefore, within sheltered instruction, the SIOP model specifically is the ideal and most advantageous method for teaching ELLs.

The SIOP model is used as an observation tool as well as a guide for planning lessons. The SIOP model is broken down into 30 components that need attention which are grouped into eight main features: lesson preparation and building background, comprehensible input, strategies, interaction, practice and application, lesson delivery, and review and assessment. Because the SIOP model is used to evaluate the fidelity of a lesson to the model, there is a rubric that is used for grading or assessing a lesson. This rubric could be used by teachers to assess their own lesson, by teachers and administrators to assess other teachers, or by professional development coordinators to assess teachers learning to implement the SIOP model. Each of the 30 components is graded on a scale from zero to four, with levels zero, two, and

four being defined explicitly (Echevarría, Vogt, & Short, 2010, p. 18). Level zero represents an aspect of the lesson that is not representative of the SIOP model at all because it does not create any supports of ELLs. Level two represents an aspect of the lesson that provides some support for ELLs but to the level that the SIOP model demands. Level two is typical of a teacher who is learning how to implement the model but has not perfected the practice. Level four, which indicates the highest level of skill, is obtained only by those teachers who have extensive training in the model and have had practice planning and carrying out the SIOP model in the classroom. By assessing lesson plans based on this rubric, teachers can improve their instruction and strive for excellence and consistency in every lesson. The consistency of implementing the SIOP model is important for the academic success of ELLs. The SIOP model should be followed exclusively because studies have shown that the amount of instruction given through the SIOP model is directly related to the success of the ELLs, with the trend demonstrating that ELLs achieve higher levels of academic achievement in districts where the level of commitment to the SIOP model is high (Echevarría, Vogt, & Short, 2010, p. xi). Because the SIOP model has been proven to be effective in the teaching of ELLs, it would be beneficial for entire districts to use the SIOP model so that high levels of academic achievement could be obtained by ELLs in across the curriculum.

In order to promote consistency of the SIOP model, extensive and continued professional development is necessary. Teachers should be involved in ongoing professional development programs rather than sporadic or short-term workshops. In an ideal situation, all content area teachers with ELLs in their classes would have SIOP training and be able to adequately differentiate instruction to make grade-level content comprehensible for ELLs. If a district chooses to use the SIOP model, all teachers that work with ELLs should be adequately trained so that they approach the education of ELLs in the same manner. New teachers should be trained in the method so that they know how to reach ELLs; in-service teachers should be trained so that they can continue to improve upon their lesson planning and implementation, and supervisors or professional development leaders must be trained in the model in order to pass the knowledge onto other educational professionals who work with ELLs (Echevarría & Short, 2000, p. 9). If all professionals were trained in the SIOP model, they

would be able to collaborate on lesson and unit planning to create an atmosphere that consistently promotes high level learning for ELLs. Unfortunately, this is not the case due to lack of time, resources, funding, and the willingness of teachers to participate in more training. Professional development programs that have been shown to be beneficial for teachers are observation, modeling, problem solving, collaboration, monitored experience, and assessment of instruction (Echevarría, Vogt, & Short, 2010, p. 10). Professional development for the SIOP model is especially useful because it can be clearly taught to teachers due to the specific guidelines and characteristics that must be included in a lesson in order for it to be considered aligned with the SIOP model.

Because the ELL population in the United States continues to grow and new research is constantly being published on the methods for educating these students, teachers should concentrate on continued professional development in order to learn about new techniques and improve open their own implementation of the SIOP model by observing other teachers and reflecting upon their own performance. In 2005, twenty-two members of the Intensive Teacher Institute (ITI), an organization established as a reaction to the lack of certified Bilingual and ESL teachers, in a high-needs district on Long Island, New York were exposed to and trained in the SIOP model and were asked to write lesson plans in accordance with the model. The participants were in-service content area teachers who did not have certification or significant training for teaching ELLs. The teachers in the ITI program were expected to choose specific aspects from the SIOP rubric to focus on and develop lesson plans that aligned with the SIOP model. The teachers then taught at least one SIOP lesson in their own classroom and observed at least one SIOP lesson of another teacher. The teachers then wrote reports describing how they implemented the SIOP model, any successes or difficulties, and the effectiveness of the model (Honigsfeld & Cohan, 2006, p. 5). After reading the reports, it was determined that the participants as a whole claimed that the SIOP model training was the most effective form of professional development that they had ever participated in because it increased their knowledge about teaching ELLs and their skill level in implementing a concrete model. To assess the lasting effects of the professional development, the twenty-two original subjects were asked to participate in an interview

with the research team and fifty percent of the original subjects agreed. The results of the interview showed that the professional development had helped the teachers learn and implement the SIOP model in their classrooms with a moderate to strong fidelity to the model.

However, due to a lack of time for planning, observation, and discussion, the SIOP lessons of these teachers were not as organized, structured, or consistent as the SIOP model is intended to be (Honigsfeld & Cohan, 2006, p. 7). These results show that professional development and training in the SIOP model does allow and encourage teachers to plan instruction based on the model. However, in order for teachers to become skillful in working with the SIOP model, professional development and assessment must be continuous rather than isolated.

As previously mentioned, BETAC provides training to those schools who request assistance. BETAC will help schools implement the model of their choosing but if the school asks for advice about how to teach ELLs, the Rochester area BETAC recommends the SIOP model. The SIOP training through BETAC consists of a two hour session once a month for one year. Observations of the schools, surveys of content area teachers, and interviews of ESL teachers within the schools during and after the training indicate that continued professional development helps teachers become more comfortable using the SIOP model and results in the consistent use of the model by the school in its entirety (J. Smith, personal communication, February 18, 2011). Looking at these findings, it is evident that attending one time workshops or reading books about the model will not be sufficient because it does not provide follow-up or ongoing support. An urban school in the Rochester area offers incentives to teachers to participate in SIOP training. If a tenured teacher completes more than 36 hours professional development or a non-tenured teacher completes more than 24 hours, they will receive a stipend (J. Brown, personal communication, March 9, 2011). By offering a stipend to teachers for obtaining a significant amount of training, the school is demonstrating its commitment to the SIOP model and the ELL population. Professional development is necessary for teachers in order to stay up to date on a topic that is constantly evolving and affects the responsibilities of many teachers in the United States.

Once teachers have been adequately trained and continue working with the model through

professional development, they are able to manipulate the model in a way that will fit the needs of the classroom and the students. The SIOP model is very flexible and can be adapted to work in any classroom situation and with students from different cultural or educational backgrounds and proficiency levels (Echevarría, Vogt, & Short, 2010, p. 20). It can be a part of other program models such as ESL programs and bilingual program as well as mainstream classes. It can be used in ESL classes when teaching language and academic processes through content to students of various proficiency levels. The SIOP model can be implemented in conjunction with a bilingual program to make English content courses more comprehensible to all of the ELLs in the class because content classes in bilingual programs are grouped by grade level rather than proficiency level. Mainstream content teachers can use the SIOP model to make content comprehensible for ELLs in an English only setting.

Because the SIOP model is simply a formal compilation of good teaching methods, content area teachers can implement the SIOP model in classes that contain English only students as well as ELLs, regardless of the ratio. The SIOP model adapts content in a way that is beneficial for ELLs, students from low income families, students with disabilities, and other students with academic disadvantages. Observations of schools that implement the SIOP model imply that the model helps average English only students grasp concepts more quickly and retain more information (J. Smith, personal communication, February 18, 2011). For ELLs in programs that lead to a transition into mainstream classes, the SIOP model helps students transition from the language program in which they were involved to mainstream classes (Echevarría & Short, 2000, p. 7). Without the SIOP model, the transition from the ESL or bilingual program to the mainstream courses would be difficult and stressful. By using the SIOP model, teachers gradually ease ELLs through the transition into mainstream classes by slowly taking away supports until they are able to actively participate in classes with English only students without assistance. In terms of material, SIOP lessons require both content and language objectives, which ensures that each lesson is planned to develop language skills and content knowledge simultaneously (Echevarría, Vogt, & Short, 2010, p. xii). The combination of language and content objectives offers infinite possibilities for various grade levels, subject areas, and skill development.

The most recent research published supports the idea that the SIOP model is the most popular and efficient instructional model for teaching ELLs. The SIOP model is a compilation for good teaching methods that increase the level of retention rate and development of academic language and content knowledge (J. Brown, personal communication, March 9, 2011). The aspects of the model that are praised most are the integration of language and content objectives, the ability to accommodate students of various proficiency levels within the same class, the identification and explanation of particular strategies to adapt content and language, the creation of authentic opportunities for meaningful communication, the establishment of classroom routines, the simultaneous development of speaking, writing, reading, and listening, and guidelines for supplemental materials for illustrating concepts such as visual aids, graphic organizers, word walls, realia, and more (Goldenberg & Coleman, 2010c, 161). All of these aspects of the SIOP model promote learning of content knowledge, procedural skills in terms of academic tasks, study skills, and learning strategies, which are essential for the success of ELLs in the academic realm. These aspects promote success because they regulate the differentiation of instruction for ELLs. Teachers who are not trained in the SIOP model often do not understand how to differentiate instruction to meet the specific linguistic needs of ELLs. Within ESL classes, differentiation of ELLs does not mean that students are grouped by proficiency level and the beginners are working on something independently while intermediate students are working with the teacher on something completely different. In content area classes, differentiation of ELLs should integrate them with English only students rather than secluding them and giving them easier work to do (J. Smith, personal communication, February 18, 2011). The SIOP model explicitly addresses differentiation of instruction for ELLs so that these students are not neglected or isolated in the classroom. The SIOP model also outlines how to remove scaffolding gradually because it will encourage students to feel comfortable in mainstream content classes. Students should be taught to recognize various supports and scaffolding used by teachers so that they can learn how to use them independently. This will also prevent students from becoming frustrated and lost when the scaffolds are no longer provided by the teacher because they will be able to use them on their own when needed (J. Brown, personal

communication, March 9, 2011). The SIOP model lays out the tools and strategies that teachers must provide and teach to ELLs in order to ensure their success in school.

Although the SIOP model is widely considered the most efficacious model for teaching ELLs, it is not the most widely used model. Most schools come up with their own philosophy of beliefs and plan instruction based on the philosophy of education for ELLs and what they believe will work for the students in their classrooms (J. Smith, personal communication, February 18, 2011). The remaining question is why school districts are not implementing the SIOP model if research shows that it is the most effective model for teaching ELLs at this time. There are various factors that affect the decision of schools to not use the SIOP model. The main issue, from which the minor issues stem, is the relatively small population of ELLs in some schools. According to Michelle Adams, the ESL teacher from rural district, she only services three Liberian ELLs; one in seventh grade, one in ninth grade, and one in eleventh grade (personal communication, March 4, 2011). Emily Davis, the ESL teacher from a suburban district, services only ten students at the secondary level (personal communication, March 24, 2011). Both teachers acknowledge that the SIOP model would be the ideal model for teaching ELLs but claim that their schools choose to not to use the model because there is not a large enough population of ELLs (M. Adams, personal communication, March 4, 2011; E. Davis, personal communication, March 28, 2011). The urban district, on the other hand, has 325 students that are identified as ELLs and the schools maintain a program that combines the use of a bilingual program and the SIOP model (J. Brown, personal communication, March 9, 2011). Commentary by interviewees and research demonstrate a direct connection between the size of the ELL population and implementation of the SIOP model.

Rural and suburban districts with small ELL populations often have a hard time understanding the importance of using the SIOP model because it affects such a small portion of the student population. It is time consuming to train all teachers in a district implement the SIOP model skillfully and consistently. Some administrators and teachers in districts with small populations of ELLs tend to think that it is not worth the trouble. Another barrier is that teachers do not have enough time to participate in training, plan lessons, and analyze their performance using the SIOP model

even if they are genuinely interested and willing to participate (J. Brown, personal communication, March 9, 2011). Secondary content teachers often find that they have trouble finding time to plan lessons for their different courses and taking care of their other obligations as teachers. Therefore, they are often unable to find time to participate in the continued professional development that would be necessary to learn how to properly use the SIOP model. Because teachers do not have the time to receive adequate training, they are not able to stay faithful enough to the SIOP model for it to be as beneficial to ELLs as it would be if it were implemented skillfully (J. Smith, personal communication, February 18, 2011). Therefore, if schools try to use the SIOP model but not all teachers are fully trained or demonstrate a complete fidelity to the model, ELLs will not progress as quickly linguistically or have as much academic success as would be expected with the use of the SIOP model. Though it is understandable that these obstacles would discourage schools from using the SIOP model, it is disappointing that schools would allow these factors to get in the way of providing ELLs with a high quality education.

The findings of a literature review and interview of professionals who work with ELLs imply that the SIOP model is the most effective

method for teaching ELLs in most situations. The bilingual method is also an effectual model but cannot be said to be the most effective because it can be used in far less situations. By analyzing the responses of the teachers interviewed concerning the education of ELLs in their schools and comparing these to literature published, connections can be made between what theorists and teachers claim is the best method and what is actually being implemented in schools. The reasons that certain districts are unable or choose not to use the SIOP model generally seem to be linked to the size of the population of ELLs. Though this correlation is valid, it is unacceptable because the primary concern of a school should be to meet the needs of all students. ELLs should not be deprived of an academically rigorous and linguistically appropriate education.

Appendix 1

Table A-5-1. Number and percentage of children ages 5-17 who spoke a language other than English at home and who spoke English with difficulty: Selected years, 1979-2008

[Numbers in millions]

Year	Total population	Number	Percent of total population	Spoke a language other than English at home		Percent of those who spoke a language other than English at home
				Spoke English with difficulty	Percent of total population	
1979	44.7	3.8	8.5	1.3	2.8	34.2
1989	42.3	5.2	12.3	1.8	4.3	34.6
1992	47.7	6.3	13.2	2.2	4.6	34.9
1995	47.5	6.7	14.1	2.4	5.2	35.8
1999	52.7	8.8	16.7	2.6	5.0	29.5
2000	52.5	9.5	18.1	2.9	5.5	30.5
2001	53.0	9.8	18.5	2.8	5.4	28.6
2002	53.0	9.8	18.5	2.8	5.3	28.6
2003	53.0	9.9	18.7	2.9	5.5	29.4
2004	52.9	9.9	18.8	2.8	5.3	27.9
2005	52.8	10.6	20.0	2.8	5.4	26.8
2006	53.4	10.8	20.3	2.8	5.2	25.4
2007	53.2	10.8	20.4	2.7	5.1	25.2
2008	53.0	10.9	20.5	2.7	5.0	24.6

NOTE: Respondents were asked whether each child in the household spoke a language other than English at home. If they answered "yes," they were asked how well each child could speak English using the following categories: "very well," "well," "not well," and "not at all." All children who were reported to speak English less than "very well" were considered to have difficulty speaking English. Spanish-language versions of both the Current Population Survey (CPS) and the American Community Survey (ACS) were available to respondents. Due to differences between the CPS and the ACS, use caution when comparing data before 2000 (CPS) with data from 2000 onward (ACS). For more information on the CPS and the ACS, see supplemental notes 2 and 3, respectively.

SOURCE: U.S. Department of Commerce, Census Bureau, Current Population Survey (CPS), 1979 and 1989 November Supplement and 1992, 1995, and 1999 October Supplement, and American Community Survey (ACS), 2000-2008.

National Center for Education Statistics (2010) *The condition of education 2010: Indicator 5 language minority school-age children* (Publication No. NCES 2010-028). Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Appendix 2

National Center for Education Statistics

Table A-5-3. Number and percentage of children ages 5–17 who spoke a language other than English at home and who spoke English with difficulty, by language spoken, region, and state: 2008

[Numbers in thousands]

Region and state	Total population	Number	Percent of total population	Spoke a language other than English at home					
				Total		Percent of total population	Percentage distribution by language spoken		
				Number	Percent of total population		Asian/Pacific Islander ¹	Other Indo-European ²	Other
United States	53,012	10,887	20.5	2,673	5.0	74.7	11.6	10.4	3.3
Northeast	9,044	1,860	20.6	396	4.4	52.1	16.9	26.7	4.3
Connecticut	604	110	18.2	17	2.8	67.3	15.3	14.6	2.8 !
Maine	202	9	4.5	‡	0.9 !	‡	‡	‡	‡
Massachusetts	1,046	211	20.2	44	4.2	50.4	22.8	23.6	3.2 !
New Hampshire	215	16	7.3	‡	1.1	‡	‡	‡	‡
New Jersey	1,488	387	26.0	71	4.8	58.3	15.8	19.8	6.2
New York	3,198	881	27.5	203	6.4	49.2	17.5	28.8	4.5
Pennsylvania	2,030	209	10.3	49	2.4	48.7	12.1	37.1	2.1 !
Rhode Island	164	33	20.2	7	4.1	77.0	9.4 !	11.5 !	‡
Vermont	96	5	4.9	‡	0.7 !	‡	‡	‡	‡
Midwest	11,605	1,336	11.5	344	3.0	63.3	12.8	17.3	6.7
Illinois	2,286	512	22.4	123	5.4	79.1	5.6	13.8	1.4 !
Indiana	1,138	95	8.4	28	2.4	61.5	8.9	29.6	‡
Iowa	505	45	8.8	10	1.9	69.3	12.6 !	17.7 !	‡
Kansas	491	58	11.9	12	2.5	82.1	9.5 !	5.6 !	2.8 !
Michigan	1,763	160	9.1	37	2.1	51.5	16.2	11.4	21.0
Minnesota	895	116	12.9	39	4.3	44.1	32.6	9.7	13.5
Missouri	1,014	64	6.3	17	1.7	45.1	19.9	28.5	6.6 !
Nebraska	327	45	13.9	13	4.0	80.3	7.8 !	5.6 !	6.3 !
North Dakota	100	7	7.0	‡	2.8 !	‡	‡	‡	‡
Ohio	1,991	119	6.0	36	1.8	41.1	10.3	35.4	13.3 !
South Dakota	145	10	7.0	‡	0.6 !	‡	‡	‡	‡
Wisconsin	950	106	11.1	26	2.7	63.5	19.2	13.5	3.8 !
South	19,628	3,483	17.8	911	4.6	83.7	6.8	7.2	2.3
Alabama	822	37	4.6	13	1.6	82.5	8.4 !	5.4 !	3.7 !
Arkansas	504	39	7.8	11	2.2	87.3	8.4 !	3.3 !	‡
Delaware	149	18	11.6	4	2.9	76.1	6.2 !	17.7 !	‡
District of Columbia	76	7	9.7	‡	1.6 !	‡	‡	‡	‡
Florida	2,871	736	25.6	149	5.2	75.3	5.0	17.3	2.4 !
Georgia	1,813	241	13.3	51	2.8	80.0	9.5	7.6	3.0 !
Kentucky	717	35	4.9	13	1.8	63.7	9.6	21.2	5.5 !
Louisiana	792	42	5.3	11	1.4	60.5	14.5 !	19.7	5.4 !
Maryland	965	135	14.0	30	3.1	45.6	19.7	28.7	6.0 !
Mississippi	554	18	3.3	5	0.8	60.7	11.4 !	8.7 !	19.3 !
North Carolina	1,595	192	12.0	62	3.9	87.7	4.8	5.9	1.6 !
Oklahoma	638	61	9.5	15	2.4	79.8	13.8 !	4.1 !	2.2 !
South Carolina	769	53	6.9	14	1.8	86.8	4.0 !	9.2	‡
Tennessee	1,055	75	7.1	19	1.8	70.5	8.2 !	8.0	13.3 !
Texas	4,722	1,609	34.1	471	10.0	92.4	5.0	1.5	1.1
Virginia	1,308	176	13.5	40	3.0	62.4	19.2	14.5	4.0 !
West Virginia	280	7	2.6	‡	0.8	‡	‡	‡	‡
West	12,734	4,208	33.0	1,022	8.0	79.2	13.5	4.8	2.6
Alaska	129	16	12.6	4	2.8	‡	41.0 !	‡	49.7
Arizona	1,203	370	30.8	103	8.5	90.3	2.4 !	1.6	5.7
California	6,666	2,929	43.9	684	10.3	80.1	14.7	4.2	1.1
Colorado	855	165	19.3	46	5.4	85.8	7.0	6.5	0.7 !
Hawaii	200	35	17.6	11	5.4	12.9 !	84.3	2.8 !	‡
Idaho	288	35	12.1	9	3.0	61.7	14.0 !	21.9 !	2.4 !
Montana	163	8	4.9	‡	1.3 !	‡	‡	‡	‡
Nevada	473	153	32.3	37	7.9	86.3	7.5	2.3 !	3.8 !
New Mexico	351	99	28.2	18	5.1	82.0	5.3 !	1.9 !	10.8 !
Oregon	627	112	17.9	29	4.7	78.9	11.3	7.5	2.3 !
Utah	577	71	12.3	17	2.9	69.8	3.4 !	6.1 !	20.7 !
Washington	1,114	210	18.8	62	5.6	63.9	19.5	12.0	4.7 !
Wyoming	89	5	5.2	‡	0.9 !	‡	‡	‡	‡

¹ Interpret data with caution (estimates are unstable).

² Reporting standards not met (too few cases).

¹ Any native language spoken that linguists classify variously as Sino-Tibetan, Austroasiatic, or Austronesian languages.

² An Indo-European language other than Spanish (e.g., French, German, Portuguese, etc.).

NOTE: Respondents were asked whether each child in the household spoke a language other than English at home. If they answered "yes," they were asked how well each child could speak English using the following categories: "very well," "well," "not well," and "not at all." All children who were reported to speak English less than "very well" were considered to have difficulty speaking English. A Spanish-language version of the American Community Survey (ACS) was available to respondents. Detail may not sum to totals because of rounding. For more information on geographic region, see supplemental note 1. For more information on the ACS, see supplemental note 3.

SOURCE: U.S. Department of Commerce, Census Bureau, American Community Survey (ACS), 2008.

National Center for Education Statistics (2010) *The condition of education 2010: Indicator 5 language minority school-age children* (Publication No. NCES 2010-028). Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Appendix 3

States with greatest ELL Student Population Growth from 1994-95 – 2004-05

State	# ELL in 2004 - 05	% Growth From 1994 - 95
South Carolina	15,396	714.2%
Kentucky	11,181	417.4%
Indiana	31,956	407.8%
North Carolina	70,288	371.7%
Tennessee	19,355	369.9%
Alabama	15,295	336.8%
Puerto Rico	578,534	304.4%
Nebraska	16,124	301.4%
Arkansas	17,384	294.6%
Georgia	50,381	291.6%
Colorado	90,391	237.7%
Nevada	72,117	208.3%
New Hampshire	3,235	198.4%
Virginia	67,933	196.1%
Delaware	5,094	183.2%
Missouri	15,403	183.0%
Utah	56,319	163.7%
Minnesota	56,829	161.4%
Iowa	14,421	148.3%
Oregon	59,908	133.1%
Kansas	23,512	131.7%
Ohio	25,518	108.4%
Wyoming	3,742	101.9%
Pennsylvania	39,847	100.3%

Sources: U.S. Department of Education's Survey of the States' Limited English Proficient Students and Available Educational Programs and Services, 1991-1992 through 2000-2001 summary reports; state publications (1998-1999 data); enrollment totals from the National Center for Educational Statistics Core of Common Data, 1998-1999.

Payán, R. M., & Nettles, M. T. (2008, January). *Current state of English-language learners in the U.S.: K-12 student populations*. Princeton, NJ: 2008 English-Language Learner Symposium.

Appendix 4

National Center for Education Statistics

Table A-5-2. Number and percentage of children ages 5–17 who spoke a language other than English at home and who spoke English with difficulty, by selected characteristics: 2008

[Numbers in thousands]

Characteristic	Total population	Number	Percent of total population	Spoke a language other than English at home				Spoke English with difficulty			
				Total		Ages 5–9		Ages 10–17			
				Number	Percent of total population	Number	Percent of population ¹	Number	Percent of population ¹		
Total	53,012	10,887	20.5	2,673	5.0	1,361	6.9	1,312	4.0		
Language spoken at home											
Spanish	7,781	7,781	100.0	1,996	25.7	1,037	34.8	959	20.0		
Other Indo-European ²	1,520	1,520	100.0	279	18.4	131	23.3	148	15.4		
Asian/Pacific Islander ³	1,155	1,155	100.0	311	26.9	152	35.2	159	22.0		
Other	431	431	100.0	87	20.2	40	23.9	47	17.9		
Race/ethnicity ⁴											
White	30,386	1,752	5.8	364	1.2	142	1.3	222	1.2		
Black	7,514	418	5.6	97	1.3	35	1.3	62	1.3		
Hispanic	10,787	7,129	66.1	1,835	17.0	998	22.9	837	13.0		
Mexican	7,622	5,180	68.0	1,447	19.0	817	26.0	630	14.1		
Puerto Rican	945	425	45.0	76	8.0	32	8.8	43	7.5		
Cuban	231	162	70.0	26	11.4	12	14.0	14	9.9		
Dominican	297	249	83.9	48	16.0	19	17.8	29	15.1		
Central American	710	568	80.0	145	20.4	74	26.5	71	16.4		
South American	453	348	76.9	59	13.0	26	15.4	33	11.6		
Other Hispanic	529	197	37.3	35	6.7	18	8.7	17	5.3		
Asian	2,063	1,323	64.1	331	16.0	165	20.3	166	13.3		
Asian Indian	389	255	65.6	43	10.9	26	14.2	17	8.1		
Chinese	435	300	69.0	82	19.0	39	22.8	43	16.5		
Filipino	339	129	37.9	33	9.8	14	11.5	19	8.8		
Japanese	58	28	49.2	12	20.2	7	31.6	4	12.6		
Korean	205	146	71.2	42	20.4	17	24.2	25	18.4		
Vietnamese	241	191	79.0	48	19.9	26	28.1	22	14.7		
Other Asian	397	274	69.2	71	18.0	36	23.2	35	14.7		
Pacific Islander	74	25	33.2	6	7.6	‡	9.2	3	6.7		
American Indian/Alaska Native	399	65	16.3	12	2.9	5	3.6	6	2.5		
Two or more races	1,624	118	7.3	18	1.1	8	1.2	10	1.0		
Citizenship											
U.S.-born citizen	50,490	8,794	17.4	1,879	3.7	1,097	5.7	782	2.5		
Naturalized U.S. citizen	530	312	58.8	64	12.1	18	13.9	46	11.5		
Non-U.S. citizen	1,992	1,781	89.4	730	36.7	246	46.1	485	33.2		
Poverty status ⁵											
Poor	8,907	2,762	31.0	886	9.9	488	13.3	398	7.6		
Near-poor	10,881	3,196	29.4	834	7.7	443	10.4	391	5.9		
Nonpoor	32,437	4,811	14.8	916	2.8	410	3.5	506	2.4		

‡ Reporting standards not met (too few cases).

¹ Percentage of the total subgroup population for that particular subgroup. For example, 3.6 percent of all American Indians/Alaska Natives ages 5–9 spoke a language other than English at home and spoke English with difficulty.

² An Indo-European language other than Spanish (e.g., French, German, Portuguese, etc.).

³ Any native language spoken that linguists classify variously as Sino-Tibetan, Austroasiatic, or Austronesian languages.

⁴ Race categories exclude persons of Hispanic ethnicity. Totals may include some racial/ethnic categories not shown separately.

⁵ Children in families whose incomes are below the poverty threshold are classified as poor; those in families with incomes at 100–199 percent of the poverty threshold are classified as near-poor, and those in families with incomes at 200 percent or more of the poverty threshold are classified as nonpoor. Detail may not sum to totals because of missing values for poverty.

NOTE: Respondents were asked whether each child in the household spoke a language other than English at home. If they answered "yes," they were asked how well each child could speak English using the following categories: "very well," "well," "not well," and "not at all." All children who were reported to speak English less than "very well" were considered to have difficulty speaking English. A Spanish-language version of the American Community Survey (ACS) was available to respondents. Detail may not sum to totals because of rounding. For more information on race/ethnicity and poverty status, see supplemental note 1. For more information on the ACS, see supplemental note 3.

SOURCE: U.S. Department of Commerce, Census Bureau, American Community Survey (ACS), 2008.

National Center for Education Statistics (2010) *The condition of education 2010: Indicator 5 language minority school-age children* (Publication No. NCES 2010-028). Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Appendix 5

NYS Language Allocation Policy REQUIRED AMOUNT OF ENGLISH LANGUAGE INSTRUCTION

Source: Section 154.2 (d) and (e) of CR Part 154
<http://www.emsc.nysesd.gov/biling/bilinged/NEWCRPT.154.html>

ESL Program					
English Proficiency Level *	GRADES K-8		GRADES 9-12		
	Number of Units		Number of Units		
	ESL ♦	ELA ♦	ESL ♦	ELA ♦	ELA ♦
Beginning	2	-	3	-	-
Intermediate	2	-	2	-	-
Advanced	1	1	1	1	1

Bilingual/ESL Program					
English Proficiency Level*	GRADES K-8			GRADES 9-12	
	Number of Units			Number of Units	
	ESL**	NLA**	ELA	ESL**	NLA**
Beginning	2	1	-	3	1
Intermediate	2	1	-	2	1
Advanced	1	1	1	1	1

One unit of study equals 180 minutes per week.

* English Proficiency Level is determined by the scores on the NYSESLAT (NYS ESL Achievement Test) or LAB-R (Language Assessment Battery – Revised)

♦ ESL instruction is provided by a teacher with an ESL teaching certification

‡ ELA instruction is provided by a teacher with ELA or K-6 teaching certification

** Students in bilingual programs must meet the ESL requirements plus one unit of study in native language arts

Commissioner's Regulations, Part 154, Laws of New York, 2007.

Appendix 6

	Conversational language tends to . . .	"Hybrid" area	Academic language tends to . . .
Shared background knowledge	... be embedded in meaningful contexts, drawing upon shared background knowledge, or existent in the moment of the conversation. When people converse, they often talk about a specific topic about which they both have at least some direct knowledge, experience, and relevant information.	While discussing unfamiliar content, adequate background knowledge is purposefully provided by a speaker. Both presentation and text may make ample use of visuals, such as charts, posters, and photos, to make the academic content more highly contextualized for the reader/listener.	... be relatively decontextualized, relying largely on information contained in the language of the oral or written text. The reader or listener has to provide his or her own relevant background knowledge or context necessary for understanding.
Contextual cues	... be fundamentally interpersonal. Face-to-face exchanges allow for contextual and interpersonal cues such as gestures, facial expressions, and intonation.	Speakers often use interpersonal cues such as gestures, speaking rate, pauses, and intonation to make the message more comprehensible.	... be fundamentally impersonal. Emphasis, mood, and tone must be communicated primarily through words and content. Face-to-face exchange is very limited or nonexistent.
Vocabulary	... use more familiar everyday words. Precise meanings generally are not as important as maintaining conversational flow and adequate mutual understanding.	Speakers and texts use unfamiliar words and expressions that make relatively simple concepts more difficult to understand.	... use specific and less familiar vocabulary that can be technical, abstract, and carry precise meanings the listener/reader is expected to understand.
Grammatical shortcuts	... frequently use grammatical shortcuts, such as "and" or pronouns, whose meanings are apparent to the participants in the context of the conversation. Talk can include nonstandard, colloquial speech.	Speakers and writers can make their content-rich message more accessible by using a conversational tone, which can include figures of speech, familiar expressions, and less density of ideas.	... frequently use specific grammatical, organizational, and presentation elements. Writing and speech is more formalized and structured. To establish authority, tone is set by an impersonal, declarative style. Language is denser with more ideas presented.
Language to express cognitive functions	... place fewer explicit cognitive demands on the reader or listener. Events and persons are known, familiar, and concrete. Inferences, analyses, and presenting reasoned arguments are typically less prominent.	Everyday conversational events can be cognitively complex and challenging, such as a child logically retelling a television episode and showing a causal sequence, or presenting an argument for why he should be able to stay out late without parents' direct supervision.	... be used for more complex cognitive functions, such as summarizing, analyzing, and explaining; relating what is read to other ideas; evaluating and critiquing arguments; composing reasoned, well-developed texts; and interpreting and solving word problems.
Exchanges	Husband and wife exchange information about what they did at work that day; each is generally familiar with what the other one does and with his or her work colleagues. In an animated conversation, friends discuss a date the night before. Two experienced divers, planning a scuba-diving trip, decide locations and gear they will need. Softball players argue over whether a player was out when she ran to first base. Family members order from a menu at a restaurant. Friends watch a movie and make occasional comments, observations, and jokes to one another. Children take turns during show and tell, sharing with the class a favorite toy from home.	An individual trying to follow a conversation about unfamiliar persons and complicated events is provided background information or shown pictures. A student converses with someone who speaks with little expression, uses little eye contact, and does not respond to others' behaviors or responses. A speaker uses expressions such as "if you juxtapose the two" instead of "if you put them side by side," or asks, "What are the parameters here?" instead of "What do we need to consider?" A person describes a television program to her friend and explains why she found it so moving and insightful.	A student reads a book or listens to a lecture on recent advances in genetics. A teacher shares an encyclopedia article on the Electoral College. A professor of contemporary literature presents a postmodern, post-structuralist analysis of <i>The Sound and the Fury</i> . A sociologist is interviewed on a news program and asked to explain the impact of the economic downturn on community-based organizations. A student must explain his reasoning, in writing, when solving a mathematical word problem. Co-workers at a public relations firm must present and argue for their proposed campaign to rehabilitate the tarnished image of a client.

Works Cited

- Coleman, R. & Goldenberg, C. (2009). What does research say about effective practices for English learners? Introduction and part I: Oral language proficiency. *Kappa Delta Pi Record*, 46(1), 10-16.
- Coleman, R & Goldenberg, C. (2010a). What does research say about effective practices for English learners? Part II: Academic Language Proficiency. *Kappa Delta Pi Record*, 46(2), 60-65.
- Coleman, R & Goldenberg, C. (2010b). What does research say about effective practices for English learners? Part III: Promoting literacy development. *Kappa Delta Pi Record*, 46(3), 106-111.
- Coleman, R & Goldenberg, C. (2010c). What does research say about effective practices for English learners? Part IV: Models for schools and districts. *Kappa Delta Pi Record*, 46(2), 156-163.
- Commissioner's Regulations, Part 154, Laws of New York, 2007.
- Cummins, J. (2001). Assessment with intervention with culturally and linguistically diverse learners. In S. R. Hurley & J. V. Tinajero (Eds.), *Literacy assessment of second language learners* (115-129). Needham Heights, MA: Allyn & Bacon.
- Echevarría, J. & Short, D. J. (2000, April). *Using multiple perspectives in observations of diverse classrooms: The sheltered instruction observation protocol (SIOP)*. Paper presented at the meeting of the American Educational Research Association, New Orleans, LA.
- Echevarría, J., Vogt, M., & Short, D. J. (2010). *Making content comprehensible for secondary English learners: The SIOP model*. Boston, MA: Allyn & Bacon.
- Escamilla, K. & Coady, M. (2001). Assessing the writing of Spanish-speaking students. In S. R. Hurley & J. V. Tinajero (Eds.), *Literacy assessment of second language learners* (43-58). Needham Heights, MA: Allyn & Bacon.
- Honigsfeld, A. & Cohan, A. (2006, April). *Lesson study meets SIOP: Linking two successful professional development models*. Paper presented at the meeting of the American Educational Research Association, San Francisco, CA.
- Mora, J. K. (2001). Effective instructional practices and assessment for literacy and biliteracy development. In S. R. Hurley & J. V. Tinajero (Eds.), *Literacy assessment of second language learners* (149-166). Needham Heights, MA: Allyn & Bacon.
- National Center for Education Statistics (1997) *Are limited English proficient (LEP) students being taught by teachers with LEP training?* (Publication No. NCES 97907). Institute of Education Sciences, U.S. Department of Education. Washington, DC.
- National Center for Education Statistics (2008) *Preparing to serve English language learner students: School districts with emerging English language learner communities* (Publication No. REL 2008049). Institute of Education Sciences, U.S. Department of Education. Washington, DC.
- National Center for Education Statistics (2010) *The condition of education 2010: Indicator 5 language minority school-age children* (Publication No. NCES 2010-028). Institute of Education Sciences, U.S. Department of Education. Washington, DC.
- New York State Education Department. (2004). *English as a Second Language: Commencement 9-12*. Albany, NY: New York State Education Department.
- No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-11-, § 115, Stat. 1425 (2008).
- Payán, R. M., & Nettles, M. T. (2008, January). *Current state of English-language learners in the U.S.: K-12 student populations*. Princeton, NJ: 2008 English-Language Learner Symposium.
- Rance-Roney, J. (2009). Best practices for adolescent ELLs. *Educational Leadership*, 66(7), 32-38.

Rothenberg, C. & Fisher, D. (2007). *Teaching English language learners: A differentiated approach*. Upper Saddle River, NJ: Pearson Prentice Hall.

Spaulding, S., Carolino, B., Amen, K. A., & Ball, K. Immigrant students and secondary school reform: Compendium of best practices. *Council of chief state school officer*, 1-81.

Tinajero, J. V., & Cabello, A. (2001). Assessing progress in second-language acquisition. In S. R. Hurley & J. V. Tinajero (Eds.), *Literacy assessment of second language learners* (27-42). Needham Heights, MA: Allyn & Bacon.

A Story of Ethnic Germans in Hungary: From their Origins to their Expulsion

Sierra Hunt

Introduction

The admiration I hold for my Oma's generation is what initially inspired me to conclude my years at Geneseo with a project about them. Some of my earliest memories are filled with my Oma's stories about her early days in Hungary. Every time I leave a morsel of food on my plate, my Oma screams "Don't you know that we had no food during the war," which then leads into an extensive recount of her early life as a German growing up in Hungary during the years of World War Two. Admittedly, my brother and I always exchanged glances and laughed – partly because such scarcity is foreign to us. As a young girl, I found these stories fascinating just by imagining my Oma as young girl, such as myself. As I got older, more and more questions began arising; the stories she told me remained as scattered fragments that never tied together in a cohesive story line.

As I interviewed my family members and researched the history surrounding their experiences, I was drawn to the complexity of the history and the profoundness of what she and her family endured; I can't never help but compare my own (once thought to be interesting, but now comparatively banal) life up to their own. When I first began this project, I was startled by the sheer lack of information present about the expulsion of ethnic Germans from Hungary. My personal connection to this piece of history sparked a small level of indignation on my own part - every aspect of history deserves, in some sense, its due attention.

I was initially intimidated, but later absorbed, by the complexity of Hungary's role in World War II; Hitler had desperately and recklessly used them as pawns in a hapless war even when defeat was inevitable for the Axis. Moreover, I was startled to learn what a central role Western Allies had in the fate of German minorities after the war. In addition, though the interviews proved difficult at times, I was provided with surprisingly informative and insightful stories. The more I researched and spoke to my relatives, the more I realized the narrow focus of my paper had to widen; the story I sought to paint broadened to include, not only the expulsion, but the history German minorities in Hungary, their experience during World War II and what happened to them in the years following the

expulsion. Finally, I was able to make sense and organize the fragmented stories I had been told about my whole life.

Preceding each chapter, I have included a section from Henry Longfellow's poem, "Evangeline.". His epic poem told of a young woman living in Acadia who, after being displaced from her homeland to new Louisiana territory, desperately searches for her true love, Gabriel. Upon first reading the poem, I was struck by the parallels to my family's expulsion; though it occurred 200 years apart and an ocean away, an extraordinary experience linked two groups of ordinary and honest people, affected by extraordinary events they had no control over. "Evangeline" speaks of the peace and contentment of Acadians as well as the beauty and fertility of the land predating English occupation; it speaks of apprehension upon foreign occupation, the sadness of leaving land they knew so well to go to a land where they were unwelcome, and the struggle of beginning again. It is for these reasons I have included relevant segments before each chapter begins, which I hope compliment all that follows in a more concise and eloquent way than I can ever manage to do.

Part 1: Before the War

"There, in the midst of its farms, reposed the Acadian village.
Strongly built were the houses, with frames of oak and of chestnut,
Such as the peasants of Normandy built in the reign of the Henries.
Thatched were the roofs, with dormer-windows; and gables projecting

Over the basement below protected and shaded the door-way.
There in the tranquil evenings of summer, when brightly the sunset
Lighted the village street, and gilded the vanes on the chimneys,
Matrons and maidens sat in snow-white caps and in kirtles
Scarlet and blue and green, with distaffs spinning the golden
Flax for the gossiping looms, whose noisy shuttles within doors"¹

The number of ethnic Germans in Hungary exceeded 600,000 just prior to World War Two.² This group of people experienced the worst of both worlds during the war: they were subject to the brutal whims of Hitler and then were rather ironically subject to punishment for Hitler's crimes. This paper seeks to shed light on ethnic German minorities in Hungary; a struggle exemplified by the stories of two families from a village in Hungary who were affected by the war, subject to a year of tense Soviet occupation, then cast out of

their homes in Hungary and forced to begin again. But perhaps this story best begins by untangling the complex web of Ethnic German origins in Hungary.

Germans arrive in Hungary

Hungary, located on the Carpathian Basin in central Europe, has been the epicenter of many turbulent wars shaking the stability of Europe. The origins of German minorities to this area resulted from three waves of immigration to present day Hungary. The third and greatest of these waves occurred in between the years of 1718 and 1787 and came to be known as “The Great Swabian Migration”³; the German peasants who migrated with this wave were called Danube Swabians in reference to the territory in Germany, Swabia, from which most (though not all) migrated.⁴

The Kingdom of Hungary had been founded sometime around 1000 AD and encompassed the present day regions of Hungary, Slovakia and Croatia, as well as parts of Romania, Ukraine, Serbia and Austria. A long period of cohesion and peace followed, but in 1526, the Ottoman Empire began placing pressure and demands placed on the Kingdom of Hungary. Soon after a hot and cold war erupted; constant conflict ensued between the two empires for nearly 150 years, with the borders of the Kingdom of Hungary remaining somewhat dynamic and undefined. The Ottomans eventually managed to take over a considerable portion of land from this Hungarian Kingdom, part of that land included present day Hungary.⁴ This Hungarian victory was fleeting. Soon after, a portion of the Ottoman territory was taken by the Austrian Empire in the late 17th century - part of the new Austrian Empire land included present day Hungarian territory.

In just one hundred years, Hungary had been shuffled between the hands of three major Empires: The Kingdom of Hungary to the Ottoman Empire to the Austrian Empire. The Habsburgs Dynasty - who had ruled the Austrian Empire at this time - forcibly expelled the Turks from the newly acquired territory. Following the Turks departure, the Hungarian land was devastated, depopulated and comprised of swamp and wilderness.⁵ The Turks had made up a large percentage of occupants, so the Habsburg emperor (Charles VI) decided to establish a decree promoting the migration of Germans to the unoccupied area. Two Habsburg Kaisers (Maria Theresa and Joseph II) further encouraged the migration of Germans to this new land between the years of 1718 and 1787.⁴ The

settlers followed the Danube River down to Hungary during the so deemed ‘Great Swabian Migration’, which ultimately lead to a two-fold increase in the number of Germans within the Carpathian Basin.⁶ Following this large wave of immigration, approximately 400 German communities⁷ were established in Yugoslavia, Romania and Hungary with 450,000 Germans⁸ in Hungary.

In Hungary, German migrants occupied four territories; one of which was the Buda environs, or the areas surrounding Budapest.⁸ Following this newly conquered land containing Germans and Hungarians, the Austrian Empire sought a compromise with the Kingdom of Hungary in 1867. The Kingdom of Hungary unsurprisingly leapt at the chance of regaining its lost homeland and the two merged, forming the Austro-Hungarian Empire. At this time, present day Hungary was a widely diverse nation: only 48.6% claimed Magyar (aboriginal Hungarian language) as their mother tongue.⁹ Like all prior times of peace, the peace enjoyed by this merger was short-lived: soon after the Austro-Hungarian Empire became the epicenter of World War I.

Following the defeat of the Central Powers after World War I the Treaty of Trianon was signed as one of the many treaties designed to force war reparations on the Austro-Hungarian Empire.¹⁰ These Treaties dramatically reestablished the post war borders of many European countries and the Treaty of Trianon specifically addressed Hungary; the Kingdom of Hungary had simply become the Republic of Hungary, consisting of 66% less land than it had before the war.¹¹ Its new neighbors included Slovakia to the north, the Ukraine and Romania to the East, Serbia and Croatia to the south and Austria to the west.

The Treaty of Trianon was established without much regard for the minority nationalities that happened to fall within the new borders¹²; many Hungarians fell outside the new Hungarian lines, while ethnic Germans in the Buda area found themselves circled within a strictly Hungarian nation.¹³ Prior to the new territorial arrangements established after World War I, the ‘problem’ of German minorities did not exist since virtually all lived in Austro-Hungary or the Bismarck Reich. Since this new border arrangement had left such a substantial amount of Germans outside their homeland, the Paris Peace Conference attempted to preclude any impending minority problems. (In all, over ten million Germans became minorities after World War I; Poland had one million,

Czechoslovakia had three million, Hungary had 500,000 and the remaining were spread between Slovakia, Romania, The Soviet Union and Serbia).² At this conference, the Minorities Treaty was established by the League of Nations and any country containing German minorities were forced to grant these minorities cultural autonomy and equal treatment. Some countries - namely Poland and Czechoslovakia - found these requirements cumbersome and generally evaded their application. Moreover, many politicians in these countries rose to power through rhetoric that portrayed German minorities as antagonists.¹⁰ In defense, German minorities frequently petitioned the League of Nations with complaints.

As a side note - German minority issues tended to be tenser in Poland and Czechoslovakia than in Hungary. Ironically enough however, Hungary was generally the most oppressive towards its minorities¹⁴ and in return, the Swabians in Hungary had their fair share of complaints against their government.⁹ Though these communities spoke German, they developed a 'Swabian' dialect that is very distinct from German proper. One of these Buda environs, located to the Southwest of Budapest, was known as Budaörs.

Budaörs

"Then came the laborers home from the field, and serenely the sun sank Down to his rest, and twilight prevailed. Anon from the belfry Softly the Angelus sounded, and over the roofs of the village Columns of pale blue smoke, like clouds of incense ascending, Rose from a hundred hearths, the homes of peace and contentment."¹

The Latozinsky family ancestors likely settled on the far eastern end of Budaörs during the Great Swabian immigration. Budaörs grew from a small town of 1000 people in the 17th century, to number over 12,000 inhabitants by the 1930's - ninety percent of whom considered themselves German descendants.¹⁵ Though it is an agrarian town, Budaörs is located only 10 kilometers southwest the major metropolitan city, Budapest. After World War I, economic conditions had deteriorated substantially throughout Germany and Hungary and by the mid 1930's, poverty levels continued to climb; agrarian towns like Budaörs were considerably affected.¹⁶ The Hungarian Republic was still in its infancy and, on top of war reparations, they now had to deal with ethnic minorities who seemed to constitute a large portion of the population.

Josef Latozinsky, born in 1898, labored as a carpenter to make ends meet for his family. He

assumed many characteristic features of a stereotypical German; he was conservative, strict and assiduous. He specialized in making coffins – a trade he learned from another carpenter in his younger years. As a younger man during World War I, his job began with bringing dead soldiers bodies back into town – an experience that scarred him deeply. If there was a just God, he once told his children, He would never let those young men die. Nonetheless, he continued with his profession, though he never again attended church. Katarina Latozinsky, born in 1897, stayed at home, helping to raise the four young children - Hans, Theresia, Maria and Stefan. Hans, born in 1925, was the oldest child and had already begun working as a baker in the late 1930's.¹⁷

On the other side of town lived the Hummels. The Hummels were, in many ways, similar to the Latozinskys – they had five children; two older boys, Paul and Luka and three younger girls, Anneta, Ines and Margaret.¹⁸ The father, Martin, supported the family through his construction business. Though they did not know each other personally, both families would be deeply affected by the War and, in time, their paths would merge.

Though money was scarce, the Latozinsky and Hummel lifestyle was characteristic of the agricultural town. A large percentage of Budaörs residents owned grape and peach farms to produce wine, but this trade produced meager wages in a country suffering from a post war economic lapse.¹⁹ Most Swabians (approximately 56 %) were engaged in agriculture in Hungary.²⁰ Whenever a needed commodity was not available in the town, the townspeople would make the 10-kilometer walk to Budapest. Their houses were small, but quaint; each had two bedrooms – one for the four children and one for the parents – as well as a kitchen. No homes in the town had plumbing, so water had to be carried daily from the town well.

Most children attended the local school run by Brothers and Nuns in the Catholic town. At that time, the Roman Catholic Church controlled what was taught in most Hungarian schools.²¹ Boys and girls were split into separate schools and all the towns' children were lumped into one class, which often numbered up to 60 children. The highest education level almost exclusively reached in the town was 6th grade, simply because this was the minimum education level compulsory by law.²¹ Most children then sought a job to help their parents economically.

The level of education attained in Hungary strongly depended upon social status. Any

education above the elementary level was reserved for elite or wealthy members of society. Following World War I, Hungary had begun to lean politically towards conservatism, which dictated a decidedly anti-liberal educational policy.²²

Rising National Identity - for Work and Bread

"Now had the season returned, when the nights grow colder and longer,
And the retreating sun the sign of the Scorpion enters."¹

The rural Swabian population did not develop an ethnic identity until the late 1800's, following a spread of Romantic Nationalism.²³ By this time, the Swabians had grown frustrated with the Magyar state; they felt it defended Jewish Hungarian Landlords at the expense of peasant farmers.²³ Fueled by this sentiment, Hungarian-German peasants came to consider themselves solely German, and the Magyar state foreign. This new nationalism also contained an anti-Semitic element; Jews were simply lumped in with the Hungarians as antagonists.²³

The Great Depression made its way across Europe in the early 1930's. The depression, coupled with war reparations, resulted in a steady and precipitous economic decline for Hungary and Germany. As a consequence, the political mood in both countries began shifting further towards the right.²² Hitler rose to power on the promise of "Work and Bread" – a promise he in fact delivered in many ways²⁴ - though his deliverance often required illegal means and were fueled by selfish reasons. He did, however, managed to gradually pull Germans out of the depths of the Great Depression and in so doing, gained immense trust from his constituents.²⁴ In Hungary, Hitler's message of economic hope and national unity seemed promising for a struggling ethnic minority. Hitler, in fact, used ethnic Germans outside of the German borders to his advantage by exaggerating their mistreatment; this tactical move on Hitler's part further entrenched German Nationalism.²⁵

After World War I, the Hungarian government included some minority representation in their government; and though a Hungarian German Cultural Association had formed in 1923, the leaders of this association were Zippers and Saxons (two other German minority groups). All in Hungary deliberately avoided Swabian representation since they were considered lower class peasants.¹⁶ The Hungarian government was particularly suspicious off this group and would not hesitate arresting any Swabians who denounced Hungary.¹⁶ Consequently, Swabians began

identifying with pan-Germanism more adamantly than other minority in Hungary.

Many in Budaörs, especially the younger generation, became radically devoted to this Nazi ideology. Pan-Germanism in these minority villages became even more solidified when radical right wing Germans began visiting these villages to spread rumors of the supposed 'wonders' awaiting them in Germany.¹⁶ The notion of belonging to a "Master Race" was understandably appealing to poverty stricken peasants. This new sense of self-worth was intoxicating, and some unquestionably accepted every aspect of the Nazi ideology.¹⁶ Some young men in these German villages were indeed promised German citizenship if they joined this movement by German agitators who visited these villages. In fact, a practice of smuggling Swabians into Germany began in the late 1930's under the misleading pretense that they would be given well-paid jobs and dormitory housing upon their arrival. They were largely unaware of the dangers of their decision²⁶; in reality, they were directly brought to SS training camps.¹⁶

In many small German-Hungarian towns, like Budaörs, day laborers would frequently get together for drinks after work and drunkenly dismiss the Magyar State while praising the efforts of Hitler.¹⁶ However, many older, more experienced men, like Josef Latozinsky, were not fooled by the brash promise of Hitler - or at least recognized that the Nazi ideology was being carried too far. Josef began seeing this Nationalism as a thin veil covering a sinister scheme. His son, Stefan, was born in 1934 and had attended a local town fair one day in 1940. He approached one stand, where he recognized a picture of Adolf Hitler being sold. He had, of course, heard of the famous Adolf Hitler and was exposed to the collective National identity spreading in Budaörs. Just as one today would buy a picture of a celebrity, Stefan bought the souvenir along with his other friends and carried it home to show his parents. Josef took one look at his son holding the picture of Hitler – a man so profound, idolized and feared. He grabbed the picture of Adolf Hitler from his son's hand and threw it in the trash.

Part 2: The Midst of WWII 1940-1944

"Alike were they free from
Fear, that reigns with the tyrant, and envy, the vice of republics.
Neither locks had they to their doors, nor bars to their windows;
But their dwellings were open as day and the hearts of the owners;
There the richest was poor, and the poorest lived in abundance."¹

Hungary's Ideological Shift

Following World War I, a democratic regime was established in Hungary. This was soon replaced by a Marxist dictatorship led by the Hungarian Communist Party. Yet again, the communist regime only lasted for a year, demonstrating the instability of Hungary at this point in time. A right wing opposition party - led by Admiral Miklos Horthy who was formerly the chief of the Austro-Hungarian Navy - overthrew this communist regime in 1919. The "Horthy Regime" then ruled Hungary until 1944. His regime was initially characterized as Moderately Conservative; however, more revolutionary, fascist-like parties existed in Hungary and Horthy's Regime began radicalizing to pull support from these right wing parties.²

The effects of the first World War coupled with the depression had a strong impact on Hungary's developing political system; jobs were hard to come by and many clung to the notion of a strong government in hopes that it would pull Hungary out of this downward spiral.³ Horthy initially proclaimed a moderate-conservative philosophy; however, the government began moving towards right radicalism when he instated the extremist, Gyula Gömbös, as Prime Minister in 1932. Gömbös had a National-Socialist vision he hoped to fulfill in Hungary, and even traveled to Berlin on numerous occasions in order to develop a political and economic relationship with Hitler. These visits ensured a close tie between Hungary and Germany.³ Gömbös' dream of seeing Hungary as a National Socialist state was cut off when he died in 1936.

In the years following Gömbös' death, political leanings strayed even farther right and Hungary began adopting anti-Semitic tendencies.³ Anti-Semitism first surfaced in Hungary as early as the 1880's.³ The cause initially grew from a resentment of the Jewish minority population, who were very successful professionals in Hungary; in fact, 50% of physicians were Jewish by 1920.³ Moreover, a disproportionately high percentage of communist leaders had been Jewish, and the coup against the Communist party provided a motive for further anti-Semitic propaganda. Horthy's regime, however, did not initially condone or permit anti-Semitism at the government level.

The most radical party in Hungary, however, was the Arrow Cross Party, led by Ferenc Szalasi, and it espoused the typical features of German National Socialism: Anti-Semitism and a belief in

the 'Master Race'.³ Szalasi was born to a German father and a Hungarian mother; he was not characterized by others as particularly eloquent or talented, but rather as a radical ideologue who relentlessly pursued his cause.⁴ He proclaimed himself a-Semitic; in other words, he believed Jews simply had no place in Hungarian society.⁵ Following Gombos' death in 1936, Szalasi inherited most of Gombos' support group.

For a time, conditions in Budaörs – and all of Hungary for that matter – appeared to improve in the eyes of the average working class. This was particularly so for the Swabians; jobs were increasing and the ethnic minorities had a newfound sense of German pride and importance.⁵ Despite the fact that Hungary's economy had improved substantially by 1937, and though Szalasi's Arrow Cross party was viewed as a serious threat, the Hungarian government system as a whole still began adopting more authoritarian aspects. The Prime Minister at the time, Daranyi, began imposing legislation against the Jewish population; Daranyi was well aware of the growing support for the Arrow Cross Party and tried appealing to more radical constituents. His legislation restricted the number of Jewish professionals to less than 20% of all Hungarian professions. In fact, he began engaging in secret negotiations with the Arrow Cross Party – much to the dismay of General Horthy. Horthy dismissed Dayani in 1938, replacing him with Bela Imredy.⁵ Imredy's political strategy, much like Dayani, was to out-radicalize the Arrow Cross Party; he passed yet another anti-Jewish legislation in his first year as Prime Minister.⁵ Horthy's patience with these disappointing Prime Ministers had worn thin; by 1939, he attempted moving the political system more towards the center when he instated the moderate Prime Minister Pal Teleki in 1939. Furthermore, he established an election to take place shortly after his instatement to have Hungary appear more Democratized.

By this time, the Arrow Cross Party was at its peak, boasting over 250,000 members. The party promised economic changes to favor farmers and working families; consequently, a sizable contingent of Swabians comprised the party.⁶ The Arrow Cross Party received 25% of the vote in 1939 – the second highest percentage in the election - but Pal Teleki gained the edge and remained Prime Minister.

Though Pal Teleki was moderate, Germany began placing pressure on Hungary in the early stages of World War Two to form an alliance;⁶

Hitler strategically offered Hungary territory that it had lost after World War I in hopes that this would persuade them into a partnership. On November 20, 1940, Teleki signed the Tripart Agreement, which established Hungary's place in the Axis alliance. In April of 1941, German soldiers began marching towards Hungary to prepare for the invasion of Yugoslavia and requested the aide of Hungary, and upon hearing this news, Teleki committed suicide; presumably, the pressure placed on him from Germany had become unbearable and despite his efforts in preserving Hungarian autonomy, he saw his efforts were futile. Soon after, the Hungarian city of Kassa was bombed, supposedly by the Soviets - though whether Germany had done so to gain Hungarian support still remains controversial.⁷ On the following day, June 27th 1941, Hungary declared war on the USSR.

Recruitment to the SS

"Alas! in the mean time

Many surmises of evil alarm the hearts of the people."
Then made answer the farmer:—"Perhaps some friendlier purpose
Brings these ships to our shores. Perhaps the harvests in England
By untimely rains or untimelier heat have been blighted,
And from our bursting barns they would feed their cattle and children."
"Not so thinketh the folk in the village," said, warmly, the blacksmith"¹

At some point during this time, even many Swabians who were not fully aware of the political happenings began sensing a radicalism of the Nazi party beyond their initial impressions. As expressed in by Balazs Szelenyi in an Ethnic conflict, "There is no direct line from German Romantic Nationalism to Auschwitz".⁸ Despite these uneasy feelings the vast majority of ethnic Germans supported Germany; it was apparent that German victory could project them to social ranks that were not attainable otherwise.

The Latozinsky family, as well as many working in Budaörs, was generally unaware of most of the ongoing governmental turmoil. By early 1942, Teresa, born in 1928, had graduated from school and began housekeeping for a local family in town. Stefan was entering his early years in school and Maria was entering the second grade. Hans, now 17 years old, had begun working for the local bakery. On the other side of town, the Hummels were also living quite peacefully in spite of the war; Paul was 18, working with his father, Luke was just 15 and had too begun working in the construction business and the three young girls were still attending grade school. In the small town, life was proceeding quite normally. It wasn't

long, however, before the Nazi rise pervaded into their small society and the political proceedings extended to their front door.

By 1943, the war in Europe had escalated, but the tide of the war began favoring the Allies after besieged German forces notoriously surrendered at the Battle of Stalingrad. Hungary had reluctantly entered into a one-sided alliance with Hitler – and upon realizing a probable Allied victory, Horthy wanted to retract this alliance. Horthy with Kallay, the Prime Minister of Hungary at the time, felt the gravity of their position in the war and began engaging in secret peace negotiations with the Allied forces. It was not long before Hitler caught wind of these secret negotiations. In March of 1944, he ordered Operation Margarethe; Horthy was invited to a meeting with Hitler in Austria while SS units were secretly sent to occupy Hungary. As Hungary quietly slipped into the hands of Hitler, Horthy was quickly losing control of his country's autonomy.⁹

At this point in 1944, Nazi forces had entered Budaörs. Once seen as allies, the German soldiers assumed a more threatening attitude towards the Hungarians, but were apparently "greeted with flowers" by many ethnic-Germans.¹⁰ Though many were unaware of the political turmoil, the more astute understood German occupation as a sign of serious contention between the once-allied countries.

By July, Hitler began to recognize the distinct possibility of failure and was growing desperate. Looking to extend the USSR's sphere of influence, Stalin sought to take over Hungary before the British and American forces invaded or before Hungary had abandoned German ties.¹⁰

With Soviets threatening to invade Hungary, Hitler decided to recruit ethnic Germans in Hungary into the German army. In 1944, the Hungarian government signed an agreement with Hitler, placing all ethnic Germans liable for military service at the disposal of the Waffen SS. Nothing so clearly demonstrates the ideological pressure exerted by the Nazis' on the Hungarians, as does the recruitment of ethnic Germans to the Waffen SS. Josef and Martin had both been recruited to the Hungarian army; their generation had trained with the Hungarians at a younger age, so they were thus incorporated into the Honved, or Home Guard¹¹ but the younger boys of the family were subject to German Waffen SS recruitment.

The Hungarian government had in fact sanctioned recruitment to the Waffen SS; they had the task of enrolling any ethnic Germans born

between 1912 and 1925 in accordance with the first act of a bilateral agreement with Germany.¹² This was initially on a volunteer basis, though strongly encouraged. In 1943, a second agreement included those aged up to 35. The final agreement was signed in April of 1944 – in this agreement, the volunteer aspect of service was abolished.

Moreover, only 10 percent of ethnic Germans were allowed to serve in the Honved;¹² however, this was later interpreted as disloyalty to the Hungarian state.¹¹ German villages were carefully combed for those available to fight. Draft notices were to be sent to German-Hungarian towns in August; males as young as 16 years old and as old as 24 years old were to serve in the German army.¹¹ German soldiers with draft notices made their way to Budaörs. One morning Hans, Paul and 16 year-old Luka received a notice from a visiting officer; in September they were to report to the Rathaus and serve in the German Army.

In September, the young men made their way to the Rathaus, or the local town hall where they were to receive instructions. Before doing so, each said a mournful goodbye to his mother. Katarina prayed for her husband, Josef, and her son's safe return. On the other side of town, Rose Hummel, with her husband in the Hungarian army and the two boys, Luka and Paul, leaving to join the German Army, was left alone to care for her three young daughters. Neither knew the fate of Germany or the fate of their families.

The Rathaus in Budaörs was located in the center of town, and all the new soldiers piled into the main room. Here, they were divided into units led by a German officer. They were hastily ushered outside Budaörs for a brief training period – lasting no more than a week, in fact, since the Soviets were fast approaching the Hungarian border.¹³ Hitler found inexperienced army units expendable and did not hesitate placing these units on the front lines against Soviet forces – Paul and Luke happened to be placed in such a unit.

Horthy was not so ready to give up his country to Hitler; he knew the Allies were likely victors and had no desire to face the Soviet army. On October 15th, 1944, it was announced that Hungary had signed an armistice with the Soviet Union. He contacted the Soviet Union and announced an official armistice. An outraged Hitler launched Operation Panzerfaust and sent Waffen-SS commandos to force Horthy to abrogate the armistice. He was removed from office¹³ and the radical Arrow Cross Party leader, Ferenc Szálasi, was placed in power. Szálasi immediately

recommenced the deportation of Hungarian Jews, which had previously been halted by Horthy (Hungary was one of the few European countries that did not allow deportation). A reign of terror was then unleashed against Jews on the streets of Budapest – a time Hans recalls acutely.

On this day of October 15th, Hans stood upon a chain bridge on the outskirts of Budapest with his army unit. The Arrow Cross Party finally gained the power they had sought for so long and were not going to waste a moment. On the opposite side of the chain bridge, Hans saw that a brigade of German soldiers had gathered up Jews and Hungarian dissenters. He watched with shock as the German soldiers lined them up, and told them to jump to their death into the Danube River and if they refused to comply, the soldiers threw them off. One by one, he saw men and women - people his own age, some older, some younger – plunging to their death for reasons that were suddenly unclear to him. 440,000 Hungarian Jews had already been sent to Auschwitz and the remaining 200,000 were now facing the brutality of the Arrow Cross Party; on this first fateful day, hundreds of Jews were killed. A friend in his unit, who had been around the same age as Hans, approached him with a shaking head. The concern in his friend's voice spoke more than his words - one day, he told Hans, they would be the ones held responsible for these actions.

His friend's words stuck him profoundly, both for their sensitive insight and startling revelation; the German 'cause' - this once appealing lure of the Master Race - was being carried to its logical end, and that end was nothing more than malevolent. For the first time, Hans recognized the gravity of the German army's actions and realized that he was living history. He continued to look on with the other men in his unit, wondering whether, in some sense, he was to blame. It was both extraordinary and terrible; what he saw would remain with him all his life.

Approach of the Soviets

"Then came the guard from the ships, and marching proudly among them
Entered the sacred portal. With loud and dissonant clangor
Echoed the sound of their brazen drums from ceiling and casement,—
Echoed a moment only, and slowly the ponderous portal
Closed, and in silence the crowd awaited the will of the soldiers."¹

The Soviets were advancing swiftly across Hungary. Hitler and Stalin both realized that Budapest was an essential political victory in the war and began giving the looming battle of

Budapest the utmost attention. Stalin had even sent a memo to the General in charge of this operation, Malinovsky, on October 28th: "The supreme command can't give you 5 days. You must understand that for political reasons we have to take Budapest as quickly as possible." Hitler, on the other hand, realized he was hanging on to the war by a thread, and this thread had become the city of Budapest.

The United States Military landing at Normandy drew forces away from the Eastern Front, so Hitler had to selectively place his panzer units. In late October of 1944, German and Hungarian forces with 70 tanks faced a Soviet unit with 627 tanks in the Great Hungarian Plain. Soviets commenced forward, out powering the Axis, and pushing farther inward towards Budapest. However, the German and Hungarian units simply could no longer stand up to the manpower of the Soviets; in the city of Debrecen (220 km east of Budapest), for example, 227 German tanks faced 773 Soviet tanks and three times as many soldiers.¹⁴

Despite the strong Soviet progression, the pace proved too slow for Stalin who was growing impatient. Stalin demanded the Soviets cross the Danube and reach the town of Budaörs by December 8th.¹⁵ The Axis counter gradually yielded to the Soviets, who continued to push forward as the Germans and Hungarians forces retreated. By November 3rd, 1944, Soviet tanks had already reached the outskirts of Budapest. On November 8th, the Soviet General Malinovsky decided to attack Budapest from the Northwest. This attack was countered by the 18th SS Panzergrenadier Division, which consisted mostly of Ethnic Germans.¹⁶ Ethnic German and Hungarian Army divisions had a relatively 'low combat value' since many soldiers crossed over to enemy side and only 1 rifle per 18 men was available. The morale of many Hungarian and Swabian soldiers was waning; food was scarce and they began questioning for what and for whom they were fighting.¹⁴

On November 23rd, Hitler demanded to the civilians that no house be abandoned without a fight.¹⁶ But in early December, Hitler realized that failure was imminent; in desperation he ordered that soldiers and recruited civilians line the streets and buildings of Budapest with explosives – an order that was not carried out due to time constraints.¹⁷ By December 20th, the Soviets had completely encircled the West and North end of Budapest; the Soviets were closing in on the encircled, combat-weary soldiers.

In the early evening of December 24th, Hans had been fighting alongside his unit in Budapest. The soldiers were then sent to the local school for shelter and sleep until the morning. Food rations were running desperately low¹⁸; by mid-December, civilians were receiving only 150g of bread per day and 120g meat on holidays. The soldiers, in particular, were fed inadequately and by the final weeks of the war, most were starving. Theresia, upon hearing where the soldiers were stationed, decided to find her brother and bring him food on Christmas Eve. Since the Soviets had encircled the Northwest portion of the city, the road that led from Budaörs to her brother's unit was clear of Soviets. She made the precarious 10 km walk into Budapest and entered the once beautiful city - now war torn and in ruins - as the artillery sounded around her. Theresia finally located the schoolhouse and found her grateful brother safely.

Theresia made it back to Budaörs just in time. When she arrived home, her mother told her the Soviets were approaching. Katarina took her children, Stefan, Theresia and Maria, to the neighbor's house where all hid in the basement cellar together. Soon after, late on Christmas Eve, the Soviets entered Budaörs.

The Soviets had completely encircled Budapest on Christmas Day, 1944. Within this encirclement were about 100,000 soldiers of the Axis. The German military authorities, fearing punishment from Hitler, placed blame on the Hungarian forces with their listless morale and frequent desertions. The Hungarian forces were split up and placed with German units to prevent further desertion.¹⁹ The two young soldiers, Luka and Paul, were placed on the Front to prevent the Soviets from entering the city.

The Siege of Budapest lasted for nearly two months, until February 13th.²⁰ The Axis forces were ordered to prevent Soviet entry at all costs. Some ethnic Germans had evacuated when the fate of the war became apparent, but the vast majority of ethnic Germans refused to leave their homes²¹; it simply did not occur to them that they would be considered culpable for Nazi atrocities, and they certainly did not expect to face repercussions by their own government. Meanwhile, the civilians of Budaörs were in the hands of the Soviets, unprotected and without further food rations. Katarina waited in the cellar for two days with her children and neighbors before daring to leave and face the Soviets.

One of the longest sieges of World War Two came to an end in early February of 1945. After

two failed attempts to relieve Budapest, the city was left to its fate.²¹ The starving and cold soldiers could no longer withstand the siege and Hungary surrendered to the Soviets. The Soviets began grouping the captured soldiers – often separating Hungarians from Germans. The Soviet Union established an order stating ethnic German soldiers were to be deported for forced labor from the occupied territories. When the deportation quotas were not met, Hungarian soldiers were deported as well. The treatment of soldiers by the Soviets was inconsistent – some were shot without provocation while others were almost invited to escape.²² Much as the treatment of soldiers, treatment of civilians was left to the discretion and whims of individual Soviets.

When a Soviet soldier approached Hans, he made sure to speak Hungarian and was grouped with other Hungarian Prisoners of War. At the time he was captured, Hans was about 30 km away from Budaörs; recognizing that his impending fate was probable to end with death or long-term imprisonment, he decided to risk escape. He managed to slip away undetected by the Soviets on guard of the prisoners. He dressed in dingy clothes in hopes that he may slip under Soviet radar as made his risky travel on the back roads to Budaörs. When Hans arrived home, the thin boy in drab garments was unrecognized by his two sisters and brother at first. When his mother finally arrived home and she saw her son sitting in the house, she scolded his foolishness for coming home when the towns were being searched for escaped soldiers.

The Prisoners of War, meanwhile, were to be taken to work camps in the Soviet Union or Poland; approximately 32,000 ethnic Germans were transported to the Soviet Union in January of 1945. To get to the railroad station, many prisoners had to march through the main road in Budaörs. As the Battalion containing many Budaörs natives marched through the town, the residents lined the streets looking for their sons, husbands and brothers. Rose Hummel lined the streets with her neighbors and three daughters and began looking for her two sons, beset with worry. Her worry deepened as more and more soldiers passed with no sign of either son; she then began calling out to the passing soldiers, asking if anyone had seen or heard of Luka and Paul Hummel. Most soldiers passing vacantly shook their heads, but one finally recognized the names being called. He turned to Rose and, his words carried the heartrending news: On a cold afternoon in the winter of 1945, Rose

found out she would never see her 16 and 19-year-old boys again.

Josef Latozinsky and Martin Hummel had fought in a Hungarian battalion and were captured after the Siege had ended. Josef was sent to a work camp in Poland, while Martin was placed on a freight train to the Soviet Union. It seemed unlikely to Martin that he would be able to return in the near future – if at all. As he sat on the train with the other prisoners, an idea crossed his mind. He wrote a note – stating his name, his town and where he was going - crumpled the paper up and threw it out the window of the train. He knew it was a long shot, but figured it was worth attempting to notify his family.

As fate would have it, a woman found it in passing one day. A short time after Martin had been taken to the Soviet Union, she personally went into Budaörs, knocked on the Hummel's door, and handed the note, a cold comfort, to Rose.

Part 3: Soviet occupation in 1945

"Then she remembered the tale she had heard of the justice of heaven;
Soothed was her troubled soul, and she peacefully slumbered till morning."¹

Under heavy Soviet influence, the Hungarian political system gradually shifted to Communism by 1945. A Provisional National Assembly was initially established, with joint power shared between various political parties; Stalin hoped to deflect Western Criticism and on the façade, Hungary appeared to give equal weight to parties across the political spectrum. In reality, however, the Soviets exerted firm control over the National Assembly; the Red Army began suppressing any other political parties propaganda activities² and when the Independent Smallholders Party beat the Communist Party in the November election, the Soviet Commissioner, Marshal Kliment Voroshilov, refused to grant the party the power they had rightly attained. Instead, a coalition government was established with Communist party leaders holding chief positions. Fulfilling Stalin's vision, the Hungarian government and the Soviet government had in many ways become one in the same.

The Siege of Budapest proved to be one of the longest and bloodiest battles of World War Two. It also marked the demise of the Nazi party - by May, Hitler was dead and Germany surrendered to the Allies. The aftermath of the war had taken a toll on Hungary: approximately 80,000 civilians and 300,000 Hungarian soldiers were killed. After a

fleeting peak in economic status, Hungary was in economic ruins once again – 25% of its industrial base was destroyed and the land given to Hungary by Hitler was lost once again. To make matters worse, Stalin refused to let any country under his influence receive aid from the Marshall Plan – a plan that allowed the rest of Western Europe to recover fairly rapidly. On top of this economic strife, over 30,000 ethnic Germans – soldiers as well as some civilians - were taken to Gulag labor camps in the Soviet Union.³

The initial days of Soviet occupation were unpredictable for those in Budaörs. Many Soviets, exhausted and embittered by the war, came in hate. Their destructive impulses and want of revenge manifested as violence against the Swabians. The soldiers were not given unequivocal orders on how to treat civilians by high authorities, and thus, many of the crimes committed resulted from an individual soldiers' disposition. Often, this disposition was hostile since most Soviet soldiers had previously been inundated with propaganda pamphlets, portraying Hungarians and Germans as barbarians.⁴ It also did not help that Budaörs had large stores of wine – the soldiers often got drunk which made their behavior more erratic and violent. Despite these predispositions, the German-Hungarians did not face the extreme inhumanity like those in Eastern Germany suffered;⁵ this is partially because many Soviets did not immediately recognize the Swabians as Germans. However, hardly a single house escaped looting; Soviet soldiers stationed throughout towns, such as Budaörs would frequently enter homes to ransack them for food – taking whatever they pleased whenever they pleased.

The new local town authorities had been appointed by the Soviet government and consisted mainly of Hungarian communists.⁵ Compulsory labor was required of some civilians – mainly ethnic Germans. In addition, most were recruited to build trenches in the areas where fighting was still occurring. The Soviets would often cruise the streets and pick up civilians for forced labor whenever work was needed. Women were often put into a field to dig trenches, so the residents would avoid walking around the town unless absolutely necessary. In fact, the overriding objective of each day had become to avoid the wrath and whims of Soviet soldiers.

One day, the Soviets needed to use the airport located in Budaörs, but feared that the airport had been lined with mines. The Soviets gathered all the women of Budaörs together and brought them to

the local airport, where the women were told to link arms and march down the runway. Under the pretense that this action was to compact the snow so the airport runway could be used, Katarina and her two daughters linked arms with their neighbors – including Rose and her daughters - and slowly marched down the airport strip. Only after did they realize the true purpose - to make sure no mines were present. Fortunately for them, no mines went off.

After Hans returned and things settled down a bit, he managed to find work as a baker for the Soviets. This fortunate profession allowed him to easy access to food that he would steal for his family. Though the Soviet soldiers were, for the most part, apathetic to those who had been soldiers during the war, the Hungarian Police were not. They had become a main concern for Hans as they persistently sought out soldiers who had fought with Germans during the war - despite the fact that this service had been compulsory. The Hungarian police would often search individual houses for unaccounted soldiers, at which time Hans would disguised himself as a woman to hide from Hungarian police.

Just a few months working in the bakeshop, an announcement was made in Budaörs demanding anyone who had served in the army to report to the Rathaus. Hans finally gave in and went to the courthouse, where the Hungarian police vigorously interrogated him. After finding out that he evaded labor camp, he was imprisoned briefly, then was forced to work for the Hungarian government for a year in Budapest. The entire concept seemed puzzling to Hans – he was informed just months earlier by the Hungarian government that he was to serve in the war, now - by a turn in politics - he was viewed by his country as an enemy, guilty of betrayal. Though this sudden shift against German minorities seems counterintuitive, the Hungarian Government desperately wanted to distance itself from the Nazi Regime; this prompted full cooperation with Soviet demands. Even in the years before World War Two, Hungary had wanted to establish a Nation State comprised of Magyar homogeneity, and this feeling intensified after the war. This mentality proved to be disastrous to the ethnic German population, who were no longer viewed as Swabian immigrants, but as unwelcome enemies.⁶ And the soldiers were their first targets.

If soldiers lived in fear of imprisonment by their government, it was nothing compared to the fear of rape Hungarian women experienced by the Soviets. This war crime was disproportionately

higher by the Soviet army than any other European army of World War Two. Soviet authorities did not generally give heed war crimes committed by Soviets, so the soldiers rarely faced repercussions. In some areas of Hungary, as many as 70 percent of women were raped⁷ – it was quickly becoming the norm, rather than the unfortunate exception. To protect themselves against the Soviets, Katarina and her daughters would sleep in groups with neighbors in one room at night. Even Hans had an odd experience; when dressing up as a woman to hide from Hungarian police, a Soviet soldier entered his house. The soldier – seeing what he presumed to be a woman alone - grabbed his leg. When Hans stood, the Soldier realized he was a man and, baffled and embarrassed, cursed him and left. Katarina and her two daughters once narrowly escaped an attack by two Soviet soldiers; they were one of the few lucky enough to evade rape. Just a few doors down from the Latosinskys, a man was shot when he stood up to the soldiers threatening to rape his wife.

Life for POW's

Meanwhile, Josef and Martin were enduring the Gulag camps. Josef was placed cleaning up a concentration camp in Poland. He happened to befriend a soldier who was running the work camp, and was sent home earlier than most. Instead of going home to Budaörs and being harassed by the Hungarian police like his son, Hans, Josef remained in Budapest. He found a job and would visit his family time to time. He rarely spoke of his experience in Poland.

The post-war USSR was in desperate need of rebuilding, so Martin worked to rebuild streets. The prisoners received dismal daily rations of one piece of bread and a bowl of hot water. If such a ration sounds like a death sentence, it was in many ways. Stalin did not care in the slightest how many died – just as long as they put in enough work before doing so. Death rates approached 25% in some camps.⁸

Martin toiled the days away, wondering about his family, his sons – whether they were all right, or whether they knew where he was. As he worked over the months, he gradually became weaker and death starting seeming imminent. After a year in the camp, he awoke one morning without enough strength to even get up. A soldier came over and kicked him, telling him to get working. When the prisoner failed to respond, the soldier gestured to another soldier, telling him to take the prisoner

away. Martin thought he would be left to die; instead, he was taken to a freight train with a group of prisoners who were also too weak to be of any use to the Soviets anymore.

It was, at first, unclear where he was being taken, but eventually Martin and the other prisoners realized they were not going to be killed; they were freed and returning home. On the way home, the train would stop periodically in fields, where the men would get off the train and rush for whatever vegetables or fruit they could find. Two young men returned to the train with raw, frozen beets, and Martin advised they not eat them; he insisted that their bodies would not be able to handle that type of food after dieting on only bread for the past year. The two young men supposed he was merely jealous - or even plotting to steal the food for himself. Martin noticed that the younger soldiers often died the quickest, making imprudent decisions in vital situations, at which point his thoughts strayed to the fate of his sons. Over the evening, dysentery took hold of the two young men and both died by the following morning.

Martin finally arrived at the train station in Budaörs. With barely enough strength to walk, he began making his way towards home. At this point, he was nearly dead and was garmented in ragged Soviet attire. In the distance, he saw his wife out in the back of the house. Rose, however, only saw what appeared to be a Russian soldier – she wondered why he was approaching their house, but was not afraid since the emaciated man hardly seemed threatening. When he reached a distance of perhaps, five feet away, Rose finally recognized her husband. Martin collapsed, and the family laid him on a bed. A doctor came to tend to the dying man: he ordered three tablespoons of soup a day to be given at first, gradually increasing the amount over time. He was only later told about the death of his two sons, when he seemed better able to handle such upsetting news. Over the course of time, Martin healed and revitalized. In vain, the Hummels, minus two sons, tried to move on from the horror of the past year.

The Latozinskys were also trying to come to terms with the extraordinary times; past times of peace seemed so vague and distant. With Hans still in a work camp, and Josef in Budapest, the future seemed uncertain and unpromising. It was only a few months longer that they would all live under Soviet rule; no one, however, expected what was to come.

Part 4: The Expulsion - January, 1946

"Soon o'er the yellow fields, in silent and mournful procession,
Came from the neighboring hamlets and farms the Acadian women,
Driving in ponderous wains their household goods to the sea-shore,
Pausing and looking back to gaze once more on their dwellings,
Ere they were shut from sight by the winding road and the woodland.
Close at their sides their children ran, and urged on the oxen,
While in their little hands they clasped some fragments of
playthings."¹

An 'Orderly and Humane' Transfer

In August of 1941, British Prime Minister, Winston Churchill met with President Franklin Roosevelt aboard a ship in the Atlantic Ocean in order to generate a list of peaceful principles. They hoped to reduce extremism and ensure civility when they officially released a declaration of peace, called the Atlantic Charter.² It is in times of war, however, that declarations are tested, and the Atlantic Charter did not live up to its own protocol when, in 1945, the Potsdam Agreement was released.

Just four years later, in July of 1945, the "Big Three" - leaders from Great Britain, United States and the USSR – met at the Cecilienhof Palace outside Berlin. The Potsdam Conference took place from July 17 to August 2 to address the preliminary issues of a war that had just previously ended. Among the agenda were the issue of war reparations, new territorial boundaries and, finally, the fate of Germans minorities who fell outside the German border. Stalin forcefully pushed for the expulsion of Germans from Czechoslovakia and Hungary. The United States and Britain were really concerned with the division of Germany, so were willing to compromise on certain issues – namely, the expulsion of ethnic Germans. Gradually Churchill and Roosevelt gave in to Stalin's demands. In an effort to avoid breach with the Soviets, the Western Allies finally agreed to the transfer of Germans in Eastern Europe.³

Article XIII of the Potsdam Agreement states, "The Three Governments having considered the question in all its aspects, recognize that the transfer to Germany of German populations, or elements thereof, remaining in Poland, Czechoslovakia and Hungary will have to be undertaken. They agreed that any transfers that take place should be effected in an orderly and humane manner."³ The reasons all three leaders finally agreed on expulsion remain rather contentious, and actually include a number of

factors. Firstly, one key factor for the Big Three was to make the Eastern European states roughly homogenous. Secondly, Stalin claimed Germans within these European countries were potential sources of future conflict, and expulsion would preclude such issues; immediately after the war, there was a distrust and sense of culpability on part of all Germans - despite what their political beliefs may have been. Thirdly, many felt the need to punish all people of German heritage based upon the premise of collective guilt; many were still coming to terms with the atrocities committed by the Nazis, and a desire to obtain justice often led to brash decisions. Finally, many propose that Stalin pushed for expulsion because he felt satellite states may look to the Soviet Union for protection against the angered Germans,⁴ and Roosevelt and Churchill simply appeased his proposals.

Hungary's inclusion in this agreement was largely due to the Soviet Union; Marshal Voroshilov urged the government to consider expelling ethnic Germans months before the conference took place. The Communists went along with Soviet opinion, driven by the notion of collective guilt,⁵ though the Hungarian Government formally renounced the principle of collective guilt upon all Germans, their sentiments and private statements often spoke to the contrary. Following the release of the Potsdam Agreement, plans began forming to organize the mass expulsion. By December, 1945, a decree was issued by the Hungarian government: Any Hungarian citizen who declared himself German, changed his last name to a "German sounding" name, or served in the SS, was to be expelled from Hungary and brought to annexed areas of Germany.⁵

Just four years prior, the Atlantic Charter had specifically stated that no territorial changes were to be made if they did not meet the wishes of those affected. But these high ideals were disregarded after the war; ethnic Germans were never asked for input and few were even aware of the decree.

Leaving Budaörs Behind

"Foremost the young men came; and, raising together their voices,
Sang they with tremulous lips a chant of the Catholic Missions:
"Sacred heart of the Saviour! O inexhaustible fountain!
Fill our hearts this day with strength and submission and patience!"
Then the old men, as they marched, and the women that stood by the
wayside
Joined in the sacred psalm, and the birds in the sunshine above them
Mingled their notes therewith, like voices of spirits departed."¹

Budaörs was the first ethnic German town targeted and the winter of 1946 was one of the

coldest in Hungary. The first train was to carry residents from the eastern section of Budaörs with very little notice or warning given. On January 29th, a Soviet soldier knocked on the Hummel's door – they had to leave their homes, pack whatever they could carry, and get to the train station a mile away within two hours. They were told that they were not going to return home, but were not told where they were being taken. Every expellee was allowed to take 100 kilograms of luggage, and every freight train consisted of 40 wagons, each carrying 30 persons.⁶ The expulsion basically extended to anyone who had committed offenses against "National Loyalty"; the fact that Hungary had sanctioned ethnic Germans joining the SS was apparently disregarded.

Martin had finally recovered from his rough ordeal during the war in the Soviet Union, and finding about the death of his two sons; the prospect of being uprooted seemed daunting. Rose took a sheet and placed it on the floor of her home. It was difficult deciding what to bring, but only practical items could be taken; the only clothes they brought were the clothes they wore. They also brought a feather blanket to battle the cold conditions. Martin suggested bringing a fork: despite their being treated as animals, they would not become animals.

They made the mile walk to the station with their neighbors to see a freight train waiting for them. Twenty to thirty people were placed in each cart, so there was very little room to move about and barely enough room for each to sleep on the floor. What is worst is that none were told where they were being taken – most however, assumed it was the USSR.

The following day, the Soviet soldiers summoned the Latozinskys. Since they were not the first in the town to be transferred, they were unsurprised - if not disillusioned - by the news. Again, they were given a couple hours to pack and get to the train station. The trains were dirty and cold and, although a train in those days could make it from Hungary to Germany in a matter of hours, it took each train twelve days.

The train carrying the Hummels stopped in the American Zone of Germany, a section in the southwest known as Huchhausen. The American General, Lucius Clay, described the disturbing sight as the first train was unloaded: "The first trainload from Hungary was a pitiful sight. The expellees had been assembled without full allowance of food and personal baggage, and arrived hungry and destitute."⁷ The Latozinskys ended up in a nearby

town, known as Obrigheim. Here, they were inspected and sterilized by American doctors.

Though the stipulations of the Potsdam Agreement urged an orderly and humane transfer, these regulations were hardly met. The Hungarian communists looted the luggage of the expellees and all arrived cold, hungry, poorly clothed and dirty.⁸ After the first few trainloads, however, these problems were sorted out and the expulsions gradually became more "humane". On June 1, 1946, the transports were temporarily halted by America because Hungary wanted to claim all ethnic German property as their own; they claimed a right to do so based on reparations for the war. Soon after, the American government refused any further expellees to the U.S. Zone of Germany since the humane and orderly stipulations of the Potsdam Agreement were not fully adhered to. In total, approximately 170,000 ethnic Germans had been transported to Germany – mainly to the state, Württemberg, though some were taken to the zone occupied by the Soviet Union. Expulsions again resumed the following year, in August of 1947, when another 50,000 ethnic Germans were displaced.⁹

Part 5: Starting Over 1946-1950

"Bearing a nation, with all its household gods, into exile.
Exile without an end, and without an example in story.

Far asunder, on separate coasts, the Acadians landed;

Scattered were they, like flakes of snow, when the wind from the northeast
Strikes aslant through the fogs that darken the Banks of Newfoundland.

Friendless, homeless, hopeless, they wandered from city to city,

From the cold lakes of the North to sultry Southern savannas,—

From the bleak shores of the sea to the lands where the Father of Waters

Seizes the hills in his hands, and drags them down to the ocean,

Deep in their sands to bury the scattered bones of the mammoth.

Friends they sought and homes; and many, despairing, heart-broken,

Asked of the earth but a grave, and no longer a friend nor a fireside.

Written their history stands on tablets of stone in the churchyards.

Long among them was seen a maiden who waited and wandered,

Lowly and meek in spirit, and patiently suffering all things."¹

The expulsion of ethnic Germans was barely known outside of Germany. However, tons of grains and potatoes were shipped from America to mitigate the food scarcity. The German economy was, unsurprisingly, in shambles after the war - industrial production had declined by 73% and unemployment was widespread.² In addition to coping with these dire post-war conditions, Germans had to somehow deal with the sudden influx of ethnic German expellees. In fact, many native Germans were forced to give up space in their homes to accommodate the refugees. The expellees found they had arrived unwelcome by the native Germans and with little prospect of work and food.

Both the Hummels and Latozinskys moved in with German families in their separate towns, Hochhausen and Obrigheim respectively. With few available jobs, food stamps supplied each family, but they never seemed to provide enough sustenance. The Latozinskys lived with the German family, and after both families acclimated and accepted their situation, they all got along reasonably well. Stefan entered school with the other children while Theresia worked in a factory, pickling food for some time. Since the Soviets claimed any German machines as war reparations, many were employed cleaning or repairing these factory machines.

Despite leaving their homeland and having to readjust to entirely new surroundings, all agreed that life was better and less dangerous in Germany than in Soviet-occupied Hungary. Stefan was called a Hungarian Gypsy by school children and bullied occasionally; otherwise, all got along well with the Germans. After Hans had been locked up a second time for being drafted in the war, he was rather resentful towards Hungary; though it was difficult starting over, he felt relieved that he no longer had to live in a country that exploited him.

Food, however, always seemed scarce and they remained entirely dependent on food stamps for years. They did find innovative ways to deal with the lack of food; trees in the town had little seeds that would fall. The younger generation figured out that when these seeds were pressed, they could make oil to be used for cooking. Many of the towns had apple orchards as well; after farmers picked the trees some apples would be left behind that were considered of poor quality to the German farmers. The children would wait for the farmers to leave, and then pick up any apples the farmers had missed.

The Hummels lived with the German family for four months; they were then able to buy a one-room place of their own on the other side of town. The Latozinskys lived with the German family for a couple of years until they were able to afford a two-room barrack that had been previously occupied by the German military. The mother and father slept in the kitchen, while the rest slept on the bedroom floor.

So much had changed for each family in such a short amount of time, but life did once again start pulling together. Each family had one another other and was free from the Soviet abuse.

Part 6: Moving on to America

In the heart of the city, they lie, unknown and unnoticed.
Daily the tides of life go ebbing and flowing beside them,
Thousands of throbbing hearts, where theirs are at rest and forever,
Thousands of aching brains, where theirs no longer are busy,
Thousands of toiling hands, where theirs have ceased from their labors,
Thousands of weary feet, where theirs have completed their journey!¹

Slowly but surely, the ethnic Germans integrated fully into German culture. In 1949, the Federal Republic of Germany was established. At this point, the expellees organized politically, developing their own 'Magna Carta.' The Charter of the Expellees was recognized on August 6, 1950 – five years after the Potsdam Agreement was signed - but was not comprised of complaints or demands; rather, it spoke of the duties of the expellees as citizens in a new country, and the rights that should duly follow.²

The two families also began integrating and becoming accustomed to their new environment. Stefan eventually graduated school and worked as a painter in a town next to Obrigheim, while Hans remained working in a factory and with the prospect of moving to America constantly on his mind. Since money was still tight, Theresia decided to venture elsewhere for work in 1948. She had heard of job opportunities in Belgium and decided to risk imprisonment and cross the border without papers. This time, the risk did not pay off when she was caught and arrested by Belgian border patrols. She was placed in prison along with other Germans who had attempted to illegally cross the border in addition to a few drug smugglers. All German citizens in the prison were sent back to Germany, but they were unsure of what to do with Theresia - a Hungarian citizen who spoke German.

After spending two weeks in a prison in Brussels, a Senator came to the prison looking, oddly enough, for a housemaid. When he saw Theresia, he asked if she would be willing to work in Belgium, taking care of his house and children. Recognizing her stroke of good fortune, she worked there for two years until she was 21, sending the money she earned back to her family in Germany.

When Theresia returned to Obrigheim, she met Johann Schefzik. Johann was also a displaced Swabian who had lived in Bayern. He had also fought in the German army and spent a stressful year evading imprisonment after the war. Johann would frequently visit Obrigheim when he went to see his cousin, Fred Hoffman. On weekend evenings, the two men would go out to drinking and dancing with other Hungarian Germans. It was

at one of these dances that Johann met Theresia Latozinsky and Fred met Annette Hummel.

Severe unemployment plagued Germany in the years following World War Two, so it was unsurprising that many refugees in the area were unable to find steady work. Due to this lack of employment, American politicians constituted a lax emigration policy³ and reserved immigration slots for some ethnic Germans – provided they found a sponsor in America. Hans and Stefan had, however, tried moving to America in 1953 when the quotas were not relaxed; as a result of the borders being closed, they were unable to attain a visa. Since they were unable to gain access to the United States, they turned their prospects north to Canada. A friend of theirs who had moved to Canada described it as a nice place and since no sponsor was required for entry into Canada, they immigrated there 1953. Maria followed her brothers soon after and all lived near each other in Canada.

Annette decided to move to America in 1950 under the assumption that economic conditions in Germany would be slow to improve. She was permitted to immigrate since a family member, her Uncle, was living in New York at the time. After a brief stay, she returned to Germany and married Fred Hoffman. Following the birth of their first daughter, they decided to move permanently to the United States, but since Annette had briefly lived in America, she was granted authorization before her husband and daughter. She left for America alone, knowing that leaving her newborn daughter would be difficult, but that doing so would put her family in a better position to immigrate over. Nine months later, Fred and their daughter journeyed to meet Annette in New York. They eventually settled down in a Queens neighborhood that was densely populated with other Germans immigrants. In time, Fred, started up a successful construction business.

After spending some time in America and taking advantage of the opportunities that were otherwise unavailable in Germany, he wrote to his cousin, Johann, telling him of all the available opportunities. Johann had been working in a fabric factory, while Theresia worked in a sewing factory; though they were living comfortably, a future in America seemed like it had more potential. Like many German expellees who immigrated to America, Johann and Theresia decided to move temporarily until things soothed over in Germany. And, like many who emigrated from Germany, they never returned.

In 1956, Johann and Theresia decided to head for America with their two-year-old son, Johnny. They found a sponsor through Johann's cousin, packed up all their belongings and, once again, left the comfort of their home for an uncertain future – only this time, it was by choice. On the boat, Theresia and Johnny slept in a small cabin with another German family while Johann slept in a separate room with thirty other men. The eleven-day ride could not have felt longer to Theresia who - like many others aboard the ship - was seasick the majority of the boat ride. When they finally reached their destination in New York harbor, they were pushed through a series of security checkpoints. After fighting their way through what seemed like endless red tape, they stepped out into the streets of New York with nothing but 300 dollars and the hope of a better life.

As time passed, each family member acquired a living and kept in close touch with one another. Josef and Katarina and Rose and Martin remained behind in Obriegheim and Hochhausen. They were getting old and starting over again in another country was simply not worthwhile. Today, Hans, Stefan and Maria live near each other in Sarnia, Ontario, and see each other quite often. Theresia and Annette live on the same block in Glendale, Queens - a block filled with many German immigrants. They remain close friends, bonded by marriage and shared experiences of their childhood in Budaörs, starting over in Germany, and raising their children in America.

Epilogue

"Still stands the forest primeval; but under the shade of its branches
Dwells another race, with other customs and language.
Only along the shore of the mournful and misty Atlantic
Linger a few Acadian peasants, whose fathers from exile
Wandered back to their native land to die in its bosom.
In the fisherman's cot the wheel and the loom are still busy;
Maidens still wear their Norman caps and their kirtles of homespun,
And by the evening fire repeat Evangeline's story.
While from its rocky caverns the deep-voiced, neighboring ocean¹
Speaks, and in accents disconsolate answers the wail of the forest"²

The expulsion of Germans from Hungary was halted in June of 1948, but the damage had already been done: the total number of Germans who had been expelled from Hungary is estimated around 180,000. In March of 1950, the Hungarian government withdrew repatriation and gave the expellees the opportunity to return to Hungary; however, only 30,000 chose to return.²

The Federation of Expellees is a nonprofit organization that was formed to represent the ethnic

Germans who were displaced from their homes after World War Two. In 2006, this group initiated the inauguration of a memorial in Budaörs to commemorate the Germans who were expelled. Stefan and his wife decided to go for the ceremony, along with other German minorities from surrounding areas.³ During the ceremony, the President of Hungary, Laszlo Solyon, apologized to all those who were affected by the expulsion.³ The president of the Hungarian Parliament, Katalin Szili, erected the memorial; engraved beneath the stone is a quote that reads in English, "Such injustice will never happen again." Stefan walked along the town he had so long ago left with a sense that a long sought for justice had been granted.

The expulsion remained largely uncovered until the fall of the Berlin wall in 1989; at this point, with the formation of the Republic of Hungary and diminishment of Soviet influence in Germany, the treatment of ethnic Germans and wrongful expulsion slowly became resurrected.⁴

It is difficult to fully appreciate what so many were forced to endure during and after World War Two. The political decisions of so few affected so many. As aptly stated by Balazs A. Szelenyi: "They started out on the eve of the First World War as privileged ethnic groups in the Austro-Hungarian Monarchy, became ethnic minorities under new nation states in the inter-war period, Übermenschen during Nazi rule, and finally ended up as refugees."⁵ What this author failed to include was that they did not *end* as refugees. They persevered, moved on, and ended up free. Integrating the hard statistics of a war with the experiences of individuals who actually lived it can allow a remote concept to be more fully understood - this cannot be illustrated more clearly by these two families.

Bibliography

Part 1: Before the War

- ¹. Longfellow, Henry Wadsworth. 1990. *Evangeline a tale of Acadie*. Champaign, Ill: Project Gutenberg: <http://www.netlibrary.com/urlapi.asp?action=summary&v=1&bookid=2009554>
- ². De Zayas, Alfred M. 1994. *A terrible revenge: the ethnic cleansing of the east European Germans, 1944-1950*. New York: St. Martin's Press, Appendix.
- ³. Prauser, Steffen, and E. A. Rees. 2004. *The expulsion of the "German" communities from Eastern Europe at the end of the Second World War*. Florence: European University Institute, p. 37.
- ⁴. Prauser and Rees. *The expulsion of the "German" communities from Eastern Europe at the end of the Second World War*, p. 35.
- ⁵. Fischer, Henry A. 2004. *Children of the Danube*. Bloomington, Ind: AuthorHouse, p. 1.
- ⁶. Prauser and Rees. *The expulsion of the "German" communities from Eastern Europe at the end of the Second World War*, p. 36.
- ⁷. Alfred De Zayas. *A terrible revenge: the ethnic cleansing of the east European Germans, 1944-1950*, p. 5.
- ⁸. Prauser and Rees. *The expulsion of the "German" communities from Eastern Europe at the end of the Second World War*, p. 38.
- ⁹. Szelenyi, B.A. 2007. "From Minority to Übermensch: The Social Roots of Ethnic Conflict in the German Diaspora of Hungary, Romania and Slovakia". *PAST AND PRESENT -OXFORD-*. (196): p. 221.
- ¹⁰. Ibid., pp. 215-216.
- ¹¹. Frucht, Richard C. 2005. *Eastern Europe an introduction to the people, lands, and culture. Volume 1*. Santa Barbara, CA: ABC-CLIO. <http://site.ebrary.com/id/10183722>: p. 534.
- ¹². Alfred De Zayas. *A terrible revenge: the ethnic cleansing of the east European Germans, 1944-1950*, pp. 14-15.
- ¹³. De Zayas, Alfred M. 1977. *Nemesis at Potsdam: the Anglo-Americans and the expulsion of the Germans : background, execution, consequences*. London: Routledge & K. Paul: p. 3.
- ¹⁴. B.A. Szelenyi, "From Minority to Übermensch: The Social Roots of Ethnic Conflict in the German Diaspora of Hungary, Romania and Slovakia", p. 247.

- ¹⁵. Budaörs, *The Innovative City – History of the Settlement*. Budaörs Official Website: <http://www.budaors.hu/en/>
- ¹⁶. B.A. Szelenyi. "From Minority to Übermensch: The Social Roots of Ethnic Conflict in the German Diaspora of Hungary, Romania and Slovakia", pp. 243-45.
- ¹⁷. Latozinsky, John. Interviewed by author. Tape recording. Sarnia, Canada., October 9th, 2010.
- ¹⁸. Klaus, Rose. Interviewed by author. Tape recording. Glendale, NY., January 11th, 2011.
- ¹⁹. Schefzik, Theresia. Interviewed by author. Tape recording. Glendale NY., January 5th, 2011.
- ²⁰. B.A. Szelenyi. "From Minority to Übermensch: The Social Roots of Ethnic Conflict in the German Diaspora of Hungary, Romania and Slovakia", p. 228.
- ²¹. Nόda Mόzes. 1970. *The Roman Catholic Denominational Education between the World Wars*. Seminar for the Interdisciplinary Research of Religions and Ideologies. http://www.jsri.ro/old/html%20version/index/no_3/noda_mozes-articol.htm.
- ²². Payne, Stanley G. 1996. *A history of fascism, 1914-1945*. London: UCL Press: pp. 268-69.
- ²³. B.A Szelenyi. "From Minority to Übermensch: The Social Roots of Ethnic Conflict in the German Diaspora of Hungary, Romania and Slovakia", pp. 230-232.
- ²⁴. Stanley Payne. *A history of fascism, 1914-1945*, p. 178.
- ²⁵. Tongue, Steven. *Hitler's Foreign Policy*. A Web Of English Histories. <http://www.historyhome.co.uk/europe/hitfor.htm>
- ²⁶. Schieder, Theodor. 1961. *Documents on the expulsion of the Germans from Eastern-Central-Europe: a selection and translation from Dokumentation der Vertreibung der Deutschen aus Ost-Mitteleuropa. 2, The fate of the Germans in Hungary*. Göttingen: Schwarz: p. 34.

Part 2: The Midst of WWII 1940-1944

- ¹. Longfellow, Henry Wadsworth. 1990. *Evangeline a tale of Acadie*. Champaign, Ill: Project Gutenberg: <http://www.netlibrary.com/urlapi.asp?action=summary&v=1&bookid=2009554>
- ². Stanley Payne. *A history of fascism, 1914-1945*, p. 131.
- ³. Ibid., pp. 268-69.
- ⁴. Ibid., pp. 271-72.
- ⁵. Ibid., pp. 274-275.
- ⁶. Ibid., pp. 276.
- ⁷. <http://www.feldgrau.com/a-hungary.html>
- ⁸. B.A Szelenyi, "From Minority to Übermensch: The Social Roots of Ethnic Conflict in the German Diaspora of Hungary, Romania and Slovakia", p. 250
- ⁹. Stanley Payne. *A history of fascism, 1914-1945*, p. 277.
- ¹⁰. Ungváry, Krisztián. 2005. *The Siege of Budapest: One Hundred Days in World War II*. New Haven, Conn: Yale University Press: p. 7.
- ¹¹. Theodor Scheider. *Documents on the expulsion of the Germans from Eastern-Central-Europe*, p. 34.
- ¹². Ibid., pp. 37.
- ¹³. Krisztián Ungváry, *The Siege of Budapest: One Hundred Days in World War II*, p. 10-13.
- ¹⁴. Ibid., pp. 18-20.
- ¹⁵. Krisztián Ungváry. *The Siege of Budapest: One Hundred Days in World War II*, p. 28.
- ¹⁶. Ibid., pp. 40-42.
- ¹⁷. Ibid., p. 45.
- ¹⁸. Ibid., p. 262.
- ¹⁹. Ibid., p. 131.

- ²⁰. Ibid., p. 195.
- ²¹. Theodor Schieder,. *Documents on the expulsion of the Germans from Eastern-Central-Europe*, pp. 38-39.
- ²². Krisztián Ungváry. *The Siege of Budapest: One Hundred Days in World War II*, p. 330.
- Part 3: Soviet occupation in 1945**
- ¹. Longfellow, Henry Wadsworth. 1990. *Evangeline a tale of Acadie*. Champaign, Ill: Project Gutenberg:
<http://www.netlibrary.com/urlapi.asp?action=summary&v=1&bookid=2009554>
 - ². Mueller, Wolfgang. 2010. "Soviet Policy, Political Parties, and the Preparation for Communist Takeovers in Hungary, Germany, and Austria, 1944-1946". *East European Politics and Societies*. 24 (1): p. 99.
 - ³. Krisztián Ungváry,. *The Siege of Budapest: One Hundred Days in World War II*, 372, p. 428.
 - ⁴. Ibid., p. 335.
 - ⁵. Theodor Schieder,. *Documents on the expulsion of the Germans from Eastern-Central-Europe*, p. 43.
 - ⁶. Ibid., p. 50.
 - ⁷. Krisztián Ungváry. *The Siege of Budapest: One Hundred Days in World War II*, p. 248.
 - ⁸. Ibid., p. 428.
- Part 4: The Expulsion January, 1946**
- ¹. Longfellow, Henry Wadsworth. 1990. *Evangeline a tale of Acadie*. Champaign, Ill: Project Gutenberg:
<http://www.netlibrary.com/urlapi.asp?action=summary&v=1&bookid=2009554>
 - ². Alfred De Zayas. *A terrible revenge: the ethnic cleansing of the east European Germans, 1944-1950*, p. 80.
 - ³. Ibid., p. 83.
- ⁴. De Zayas. *Nemesis at Potsdam: the Anglo-Americans and the expulsion*, pp. 25-30.
- ⁵. Theodor Schieder. *Documents on the expulsion of the Germans from Eastern-Central-Europe*, pp. 61-62.
- ⁶. Ibid., p. 63.
- ⁷. Alfred De Zayas. *A terrible revenge: the ethnic cleansing of the east European Germans, 1944-1950*, p. 114.
- ⁸. Theodor Schieder. *Documents on the expulsion of the Germans from Eastern-Central-Europe*, p. 64.
- ⁹. Ibid., p. 66.
- Part 5: Starting Over 1946-1950**
- ¹. Longfellow, Henry Wadsworth. 1990. *Evangeline a tale of Acadie*. Champaign, Ill: Project Gutenberg:
<http://www.netlibrary.com/urlapi.asp?action=summary&v=1&bookid=2009554>
 - ². Alfred De Zayas. *A terrible revenge: the ethnic cleansing of the east European Germans, 1944-1950*, p. 126.
- Part 6: Moving Forward**
- ¹. Longfellow, Henry Wadsworth. 1990. *Evangeline a tale of Acadie*. Champaign, Ill: Project Gutenberg:
<http://www.netlibrary.com/urlapi.asp?action=summary&v=1&bookid=2009554>
 - ². Alfred De Zayas. *A terrible revenge: the ethnic cleansing of the east European Germans, 1944-1950*, p. 130.
 - ³. Ibid., p. 126.
- Epilogue**
- ¹. Longfellow, Henry Wadsworth. 1990. *Evangeline a tale of Acadie*. Champaign, Ill: Project Gutenberg:
<http://www.netlibrary.com/urlapi.asp?action=summary&v=1&bookid=2009554>

- ². Theodor Schieder,. *Documents on the expulsion of the Germans from Eastern-Central-Europe*, 70.
- ³. Dr. Gehl, Hans; translated by Tullius, Nick. Donautal-Magazin Jg. 30, Nr. 140 dated 1. *Memorial to the Expulsion of Germans in Budaörs*. August 2006: 8-9. <http://www.dvh.org/hu-highlands/ofner-bergland-budapest/mem-germans-budaors.htm>
- ⁴. Prauser and Rees. *The expulsion of the "German" communities from Eastern Europe at the end of the Second World War*, p. 37.
- ⁵. B.A Szelenyi. "From Minority to Übermensch: The Social Roots of Ethnic Conflict in the German Diaspora of Hungary, Romania and Slovakia", p. 251.

Government-Taxpayer Tensions in the Wake of FIN 48 and Schedule UTP

Kevin Muller

Introduction

The Corporate Income Tax

Since the adoption of the United States Constitution endowing Congress with the power to “...lay and collect taxes, duties, imposts, and excises, pay the Debts and provide for the common Defense and general Welfare of the United States” in 1789 (art. I, §8, cl. 1), spirited and evolutionary debate has focused on the manners in which the government should collect revenues and monitor citizens’ compliance with those manners. Though the underlying purpose of the U.S. taxation system is of course to collect revenue for use in government-directed programs, the specific sources of wealth that have been subject to – or not subject to – taxation have been chosen in part to encourage social, economic and political initiatives of national and state governments with due consideration toward the fairness of the system itself and the economic burden that different taxes place on various classes of citizens. Today, taxpayers are subject to a complex, nuanced and unwieldy tax code: According to a report released on Aug. 27, 2010 by the President’s Economic Recovery Advisory Board, American taxpayers spend 7.6 billion hours and roughly \$140 billion a year to comply with federal tax regulations (Montgomery).

Until the passage of the Revenue Act of 1861, government revenues were collected exclusively through tariffs, sales and property taxes (U.S. Treasury 2010). In 1862, the government first levied a tax on personal income using a system conceptually not far removed from the one used today. Employers withheld taxes, deductions were permitted and higher incomes were taxed at a greater rate than lower incomes. The U.S. Supreme Court ruled such a system unconstitutional in 1895 (*Pollock v. Farmers’ Loan & Trust Co.*) on the basis that certain of these taxes could be classified as “direct,” thereby violating Article I of the U.S. Constitution: “Representatives and direct Taxes shall be apportioned among the several States which may be included within this Union, according to their respective Numbers...” (art. I, §2, cl. 3).

The taxation of corporations began in 1909 with the passage of the corporate excise tax as part

of the Payne-Aldrich Tariff Act. The act applied definitions of taxable income and tax rates to the corporation without regard to the status of its owners.

The corporate excise tax was able to come to fruition in advance of the Sixteenth Amendment because it was considered to be an excise, indirect tax and thus not prohibited by Article 1, §9 of the Constitution. In the early part of the twentieth century, most corporations distributed virtually all of their profits as dividends, and so the corporate tax at its inception primarily took the form of withholdings on dividends. In addition to being a source of revenue when individual taxes were disallowed, the act of taxing corporations instead of individuals had administrative benefits: “Because of its regular and open distribution of dividends, the corporation was an obvious target for an expansion of stoppage-at-the-source collection efforts that had proven so successful during the Civil War and Reconstruction. In effect, the corporate tax was thought to be a necessary mechanism for enforcing a comprehensive scheme of individual income taxation” (Bank, 452). The principal method of enforcing personal income tax compliance during the era had been to require taxpayers to sign an oath swearing to the accuracy of the return (Bank, 519); tax assessors were usually elected by the local taxpayers and did not always act in the interest of the government, so evasion was rampant. Because corporate dividends were often distributed regularly, formally and publicly according to state corporation laws, withholding taxes on their distribution provided a convenient method of collecting revenue in such a way that the unpopular and controversial inquisition into the private matters of individuals was not necessary.

Today, the relative administrative advantages of collecting taxes from corporations are less compelling than they were a century ago. Information technology advances have made employee withholdings automated. Moreover, corporate shares are increasingly owned by pension funds, mutual funds and other corporations and so the tracking of dividend payments is not as straightforward as it was in the days of paper stock certificates. Reporting requirements already provide the IRS with data concerning wages, withholdings, annuity and pension payments,

Social Security benefits and a variety of other transactions affecting income and potential deductions. Door-to-door surveys have been replaced with dual-reporting systems and electronic filing software that improve drastically the ability of the IRS to verify information reported on individual tax returns without personal inquisition.

Though discourse commonly treats the corporation – and governments, for that matter – in the abstract as an entity separable from the sum of its parts, corporations are better characterized as a vehicle through which individuals can organize in order to operate more conveniently and efficiently than they would as a collection of legally distinct parties. In 1894, Sen. Anthony Higgins (R-Del.) defined corporations as “but aggregations of the capital of individuals for joint profit, with joint liability and joint loss, conveniently divided into shares for the purposes of distribution and management” (Higgins). Though partnerships also represent a form of organization for business purposes, partnerships are a pass-through entity for taxation purposes; income, credits and deductions are allocated to the individuals that share ownership and are taxed according to the status of those individuals. Sole proprietorships are taxed at the individual level as well, but current tax reporting requirements make it difficult to statistically determine the breakdown of revenues between sole proprietorships and individuals.

In 1913, Congress passed the Sixteenth Amendment, which gave the government the authority to collect income taxes without regard to the census of individual states. Until 1936, dividends were excluded from personal income and so corporate income was taxed only once; in 1936 that exclusion was removed. So began the “considerable tension between the corporate income tax and the individual income tax, because corporations are owned, directly or indirectly, by individuals who (ultimately) receive a share of the corporations’ incomes” (IRS Data Release, 284). Much of this tension has centered on the so-called “double taxation” of corporate income: Revenues are taxed once as income to the corporation - at applicable corporate tax rates, which may, depending on the taxpayer, be substantially greater than corresponding individual rates and dividend/capital gains rates – and again when received by individual owners as dividends or capital gains.

Upon passage of the Payne-Aldrich Tariff Act in 1909, President William Taft defended a corporate tax because it served as “an excise tax

upon the privilege of doing business as an artificial entity and of freedom from a general partnership liability enjoyed by those who own the stock” (Taft). An official report at the time countered that corporations represented “normal and necessary forms for doing business … that hence in creating corporations a State should be considered as performing a duty rather than granting a privilege. [T]he property and the business of corporations … should be taxed with no exceptional machinery” (Report).

Today, corporate taxes represent an important, but relatively minor source of federal funds; in 2008, they provided 12 percent of total federal revenues (Tax Policy Center). Imposing a corporate tax allows the government to grant preferential treatment to certain organizations by classifying them as non-profit or creating tax incentives that affect targeted industries or types of companies. A more frequently cited argument for continuing to tax corporations, though, is that absent corporate taxes, the corporate form of organization itself might be used as a tax shelter for individuals. That is, individuals could shift part or all of their wealth into corporations and thereby enjoy a lower tax liability than wage employees. “As long as dividends were taxed as ordinary income and the accumulated earnings tax was strict enough, it was difficult to use the corporate form to shelter a great deal of income [prior to the corporate relief package of 2003 that taxes dividends at a marginal rate of 15% instead of at the taxpayer’s ordinary rate],” wrote senior specialists in economic policy and public finance in a 2007 report to Congress (Gravelle, Hungerford 2007).

The argument that taxing corporations serves as a backstop for wealthy investors is based on the presumption that the corporate tax is paid wholly by shareholders, but some economists argue that this is not the case. In the phrasing of the Congressional Budget Office, “A corporation may write its check to the Internal Revenue Service for payment of the corporate income tax, but that money must come from somewhere: from reduced returns to investors in the company, lower wages to its workers, or higher prices that consumers pay for the products the company produces” (CBO 1996, 2). Historian Bruce Bartlett observed, “The [corporate] tax remains a major source of revenue for governments at all levels and periodically draws the ire of tax reformers, who feel that corporations are not paying their ‘fair share’ of taxes – based on a naive and incorrect assumption that, if corporations paid more, other Americans would

pay less" (Bartlett 1985, 1). Bartlett argues that the double taxation of corporate income also has a significant effect on the cost of equity financing because taxes on dividends reduce increase the cost of investing; this distortion may contribute to the decision to raise capital through debt instead of equity financing and incur interest expense.

Joel Slemrod, a professor of economics at the University of Michigan, argues that the double taxation of corporate income is inefficient from a policy perspective (Slemrod 2004). He states that dividend payouts are easier to monitor and tax than corporate income and that corporations might be less likely to aggressively pursue tax shelters if tax savings accrue directly to shareholders and not to the corporation itself (21).

Determining the true incidence of the corporate income tax is further complicated when one considers the responses of each group to the effects of the tax and the residual effects of those responses. Existing literature regarding who bears the incidence of corporate taxes is varied and does not point decisively to one conclusion over another (CBO 1996, 30). As recently as December 2010, the CBO has written that "households bear the burden of corporate income taxes, but the extent to which they bear that burden as owners of capital, workers, or consumers is not clear" (CBO 2010, 2).

Accounting for Income Taxes

Though the very first corporate taxes were exacted based on income as defined by accounting standards, it quickly became clear that the goals of accounting systems were divergent from those of taxation systems and so tax laws were developed that included "specific definitions of many items of income and deductions, and many pages specifying when and how to account for the items..." (IRS Data Release, 285).

Though taxes are not calculated using financial reporting standards, corporations have to record the cash flow and economic effects of income taxes using the framework of generally accepted accounting principles. Because GAAP uses an accrual basis of accounting where revenues and expenses are recognized when they are incurred and not when cash is received or remitted (SFAC No. 1, 3), a corporation's recorded income tax expense can frequently be markedly different from the sum of income tax liabilities reported to various regulatory agencies for a given period. Additionally, the range of items that qualify as deductions for income tax purposes differ from the

range of items that qualify as expenses for GAAP reporting, so certain permanent differences between "book" and tax income exist that need never be reconciled. The manner through which companies have been required to communicate information about their obligation to pay taxes has evolved over the past century; currently, companies record a "current" liability or asset that corresponds to the cash amount of taxes to be paid or refunded, and a "deferred" liability or asset representing future taxable income or future deductions from income that have been recorded in the company's GAAP financial statements but will not be considered a taxable event until a future period (ASC 740-10-25-2).

For example, corporations are permitted to use accelerated cost recovery methods for tax purposes that allocate depreciation deductions/expense differently than methods typically followed under GAAP. While the sum total of depreciation for a given asset will ultimately be the same under both methods when all years are considered, the depreciation for any individual year will likely be different under each system. Other areas where temporary differences frequently occur include the provision for bad debts, pension contributions, unearned revenue and the carry-forward or carry-back effects of tax losses. Substantial book-tax differences can arise for companies with subsidiaries, since the IRS uses different consolidation rules than GAAP. Temporary and permanent differences are currently outlined to the IRS using Schedule M-3.

Publicly-traded corporations have an obligation to shareholders to maximize income and create wealth. Reducing tax expense has an immediate effect on bottom-line profitability and corporations therefore have an interest in minimizing tax expense. In the short term, it is possible for corporations to keep taxable income low while keeping reported net income high through the careful use of accounting techniques that defer taxable income and take advantage of preferential tax treatments. Research suggests, though, that firms cannot permanently sustain large book-tax differences without being aggressive and potentially risking IRS adjustments upon audit (Mills 1998, 355).

Many corporations are proactive in making business decisions that avoid the realization of taxable income. These corporations create tax savings by shifting income into low-tax jurisdictions and take advantage of existing deductions, tax credits and organizational structures

that entail a preferential tax status. The term “tax planning” can refer to the collective effort of management to consciously minimize tax expense through elective decisions. Though firms of all kinds – as well as individuals – can engage in tax planning at some level, a 2009 study using data from 1998 finds that the companies which manage and measure their tax departments as profit centers versus firms that view their tax department primarily as a cost center tend to be larger and often yield lower effective tax rates and higher coordination between departments (Robinson et. al. 2009, 24-25). A corporation’s effective tax rate is most often calculated by dividing accrued income tax expense by pretax income from continuing operations.

Most firms are interested primarily in reducing accrual, not cash, tax expense (Robinson et. al. 2009, 30). A possible reason for this behavior is that executives tend to be compensated based on after-tax GAAP earnings. Additionally, because the cash payment for taxes for any given tax year is not publicly available, accrual tax expense tends to be the most convenient measurement of a firm’s tax performance for investors. Though cash outflows related to taxes are separately identified in the statement of cash flows, this number may represent years of back taxes or prepayments in addition to the current year tax bill.

Appendix B illustrates selected tax information of General Electric Company for 2008-2010. GE’s reported 2010 current tax expense of \$4 million on pretax earnings of \$14,208 million drew much attention in popular media, but the situation is not as simple as it might seem. First off, the accrual-based tax provision reported in the company’s income statement for 2010 was \$1,050 million – still a fairly paltry effective tax rate of 7.39%, but a far cry from just \$4 million. An examination of the notes to GE’s financial statements reveals that though the company is subject to the top U.S. marginal corporate tax rate of 35.0%, the tax effects of global operations reduced this rate by 25.8%. Roughly 64.3% of GE’s 2010 net earnings were derived from international operations, and a substantial portion of these earnings were reinvested indefinitely and therefore subject to tax rates of less than 35%. U.S. tax law currently defers the imposition of U.S. tax on certain active financial services income until that income is repatriated to the U.S. as a dividend. Required financial statement notes related to income taxes are not so detailed as to allow a recreation of a company’s specific transactions that

yield current and deferred tax expenses and benefits, but GE’s statements demonstrate that accrual techniques and international tax laws can have a dramatic impact on a company’s current tax expense and therefore a more comprehensive understanding of tax expense is required to evaluate the extent to which a company effectively manages its tax expense.

In order to realize actual net tax savings, firms must consider both the tax and economic effects of a decision. For example, a firm might enjoy a lower marginal income tax rate by operating out of a certain jurisdiction, but the cost of constructing a plant and transferring resources to that jurisdiction might well outweigh any pure tax savings. In its Guiding Principles of Good Tax Policy (Appendix A), the Tax Division of the American Institute of Certified Public Accountants writes, “The effect of the tax law on a taxpayer’s decisions as to how to carry out a particular transaction or whether to engage in a transaction should be kept to a minimum ... That is, taxpayers should not be unduly encouraged or discouraged from engaging in certain activities or taking certain courses of action primarily due to the effect of the tax law on the activity or action.” The AICPA acknowledges, though, challenges that “stem from the desire to use the tax law for more than raising revenue, for instance, to implement social or economic policies” and notes that tradeoffs must be made among all the considerations of a fair tax system in order to create an optimal model (AICPA, 13).

Tax Avoidance versus Tax Evasion

To fully appreciate the complexities that make requirements like Financial Accounting Standards Board Interpretation No. 48 and the Schedule of Uncertain Tax Positions unpopular among some taxpayers, one must consider the difference between tax avoidance and tax evasion, as well as the difficulty of distinguishing between the two. Tax avoidance refers to strategically making business decisions that are fully legal but take advantage of preferential tax jurisdictions and treatments. For example, an individual might avoid incurring taxes on income by investing in tax-free municipal bonds, instructing an employer to deposit wages into a tax-free individual retirement account or choosing to realize capital gains only when he can realize losses to offset the taxable gains. Tax evasion, on the other hand, generally refers to illegally and knowingly underpaying taxes either by intentionally reporting taxable income incorrectly

or incompletely or by engaging in transactions that create tax losses but bear little or no economic risk. In the former case, issues of legality are usually cut-and-dry; taking a job that pays wages “under the table” and not reporting that income is clearly tax evasion. Classifying transactions that might be considered tax evasion on the basis that they reduce tax liability outside the spirit of the law, however, can involve a substantial level of judgment, especially when those transactions are made at the corporate level and involve complicated legal and financial steps.

In some cases, corporations and in particular large, international firms with tax haven subsidiaries (Lisowsky 2010, 1711) are able to take advantage of tax avoidance transactions, defined by the U.S. Treasury as “any transaction in which the reasonably expected pre-tax profit … of the transaction are insignificant relative to the reasonably expected net tax benefits … of such transaction” (Treasury 1999, 157). The IRS characterizes these abusive shelters as “very complicated transactions that sophisticated tax professionals promote to corporations and wealthy individuals, exploiting tax loopholes and reaping large and unintended tax benefits” (GAO Testimony 2003). Nevertheless, “Under present law, there is no uniform standard as to what constitutes a tax shelter; however, there are statutory provisions, judicial doctrines, and administrative guidance that attempt to limit or identify transactions in which a significant purpose is the avoidance of evasion of income tax” (Joint Committee on Taxation 2002, 4).

Since many abusive shelters are based on legal tax planning techniques, it is often difficult for courts to distinguish legitimate tax planning efforts from illicit shelters (Levinsohn 2005). The Supreme Court has repeatedly affirmed that all taxpayers retain the right to reduce their tax liability or avoid paying taxes altogether provided they do so within the boundaries of the law. Frequently, though, courts defer to the economic substance doctrine frequently credited to the 1934 *Gregory v. Helvering*. In this and many cases since, courts have found that tax savings based on literal and straightforward application of tax law are not permissible if the taxpayer realizes no economic benefits from entering into a transaction aside from the tax savings themselves. The IRS has more recently clarified that transactions may be legal even if they do not actually yield a profit as long as a cash flow or net present value analysis proves that the taxpayer could have reasonably realized a profit

(IRS Notice 2002-50, 6-B-III). The IRS also advocates employing the substance over form test, that is, a set of transactions yielding the same end result should be treated in the same way for tax purposes regardless of the legal steps used to achieve that end (IRS 2002-50, 6-C). Nevertheless, federal courts have sometimes ruled that transactions are allowable even if executed exclusively for the purpose of realizing tax savings.

Often, tax avoidance transactions are created and sold to corporations and individuals as “tax shelters.” The U.S. Senate Permanent Subcommittee on Investigations found that between 1997 and 2001, KPMG sold four “active tax products” to more than 350 individuals which generated revenues for the firm in excess of \$124 million; three of those products were later determined by the IRS to be potentially abusive or illegal tax shelters (U.S. Senate 2003, 27). In that 2003 report, the Committee concluded that “the tax shelter industry as a whole remains active, developing new products, marketing dubious tax shelters to numerous individuals and corporations, and continuing to wrongfully deny the U.S. Treasury billions of dollars in revenues, leaving average U.S. taxpayers to make up the difference” (3). In a tax shelter case involving global accounting firm KPMG, internal emails sent by tax professionals demonstrated calculations showing that the penalties for not registering tax products with the IRS, thus violating federal tax shelter laws, would be low compared to the fees generated by selling the products to clients, and that registering the tax shelters with the IRS would place the firm at a competitive disadvantage (13).

Though methodological and data constraints make it difficult to estimate with certainty the dollar amount of potential tax revenues lost through the use of tax shelters, an IRS contractor estimated that the average between 1993 and 1999 was between \$11.6 billion and \$15.1 billion per year (GAO Testimony 2003, 6). The U.S. Treasury has expressed concern about the proliferation of corporate tax shelters not only because of the short-term reduction in the corporate tax base but also because “corporate tax shelters breed disrespect for the tax system – both by the people who participate in the tax shelter market and by others who perceive unfairness” (U.S. Treasury 1999, iv). Additionally, the costs incurred by users and promoters of tax shelters to create and defend tax benefits are economically unproductive. For all of these reasons, the government has in the past

decade especially taken a serious interest in reducing the prevalence of tax shelters.

Investors, too, can easily justify an interest in understanding how aggressive a particular company is in avoiding taxes and utilizing tax shelters. Public exposure of a company's use of tax shelters is almost certainly a negative event – often, penalties and legal defense expenses are incurred, and there exists a public characterization of firms that evade taxes as poor corporate citizens. Some investors may also view aggressive tax behavior as an indicator of poor corporate governance and wonder if the firm is cheating not only the IRS but investors as well (Hanlon, Slemrod 2006, 12). On average, news of a corporation being involved in a tax shelter negatively affects that firm's stock price, especially for retail firms (Hanlon, Slemrod 2006, 5). Hanlon and Slemrod note, though, the difficulty of gauging the actual or perceived earnings effect of a tax shelter savings reversal since it is nearly impossible to reliably determine when a firm booked the savings initially and what amount of savings have been surrendered after exposure of the tax shelter since any valuation reserves would have been recorded in aggregate (28). Hanlon and Slemrod did not find a significant correlation between stock prices and the release of reports from Citizens for Tax Justice identifying firms that pay little or no taxes (35).

Controversies Arising from Disclosing Uncertainty

An individual tax position may involve a deduction or exclusion from income, the claim of a tax credit, the use of a carry-forward or carry-back loss, the classification of an entity, transfer pricing allocation or a number of other assertions made on a tax return for the purpose of minimizing income tax liability. Because modern tax law is complex and does not specifically address every conceivable transaction structure or situation but, like GAAP, institutes rules that companies must apply to their respective industries and circumstances, managers must make a conscious decision to report or not report certain events in a tax return where ambiguity exists regarding the specific application of regulations.

In March 1975, the Financial Accounting Standards Board issued Statement of Financial Accounting Standards No. 5, which requires firms to accrue a contingent liability such as income taxes payable if such liability is "probable" and "the amount of the loss can be reasonably estimated" (FAS 5, 8a-b). This guidance allowed substantial flexibility for firms to gauge materiality in their

decision to accrue or disclose, and firms developed divergent policies for dealing with SFAS No. 5 based on a variety of factors. Research indicates that many firms under-disclosed material claims and did not provide the detailed information required by SFAS No. 5 (Gleason and Mills 338). In June 2006, the Financial Accounting Standards Board, a nongovernmental body that develops accounting techniques for U.S. companies, issued an interpretation of SFAS No. 109, which had established the asset-liability approach to recording temporary and permanent book-tax differences. FASB Interpretation No. 48, now a part of Accounting Standards Codification 740-10, "prescribes a recognition threshold and measurement attribute for the financial statement recognition and measurement of a tax position taken or expected to be taken in a tax return" (FIN 48, Summary).

While FIN 48 established a needed consistency for treating contingencies related to UTPs, it contains provisions that depart from the foundational goal of financial reporting – providing relevant, reliable information to third-party users of investors and creditors – in favor of an idealistic model that assumes companies will remit all possible taxes to all possible collecting agencies even when the practical likelihood of such an occurrence is remote. FIN 48 caters to another third-party user that stands to benefit from the profitability of corporations: the government, and by extension, all who benefit from the expenditures of government. In effect, FIN 48 asks corporations to audit themselves for tax compliance and then publicize a hypothetical income tax liability that is likely different from the one the company has proposed to the Internal Revenue Service under the voluntary reporting system addressed in *Flora v. United States*; in that case, the Supreme Court noted that "our system of taxation is based upon voluntary assessment and payment, not upon restraint" (362 U.S. 167) Normally, both individual and corporate taxpayers are permitted to assess their own tax liability using the instructional tools made available by the IRS and state and local governments with the understanding that their self-evaluation may be audited for completeness and accuracy and that penalties, interest and potentially criminal sanctions can be imposed if the original return is found to underreport tax liability and sufficient disclosure has not provided to apprise the regulatory agency of the taxpayer's uncertainty in taking the positions in question. Though the public has always had a justified interest in preventing

outright tax evasion, FIN 48 potentially serves as a disincentive to corporations seeking to implement tax savings strategies in that it exposes those strategies to regulators who are interested in collecting maximal revenue.

Beginning in 2011, companies will face even further disclosure requirements: The largest U.S. corporations will be required to submit a completed Schedule of Uncertain Tax Positions (Form 1120) to the IRS. Schedule UTP requires companies to list all uncertain tax positions being taken in that year's return as well as in prior years' returns from 2010 onward and, for each listed UTP, to indicate whether the UTP represents a permanent or temporary difference and whether the UTP represents a "major" position, defined as 10 percent or greater of the total amount of UTP reserves. Corporations are also required to rank the UTPs based on the size of the associated reserves. Although the requirements of Schedule UTP for the most part build upon the analysis prescribed by FIN 48, the instructions also indicate that a taxpayer "must report on Schedule UTP a tax position taken on its return for which no reserve for income tax is recorded if the tax position is one which the corporation or a related party determines the probability of settling with the IRS to be less than 50 percent and, under applicable accounting standards, no reserve was recorded in the audited financial statements because the corporation intends to litigate the tax position and has determined that it is more likely than not to prevail on the merits in the litigation" (Schedule UTP, Instructions). That is, taxpayers must identify to the IRS any position that is likely to be technically objectionable even if they, their legal counsel and their independent auditors believe the position will likely be sustained in court.

FIN 48 and Schedule UTP have substantially heightened disclosure requirements related to corporate income taxes and the effects of each on individual corporations are contingent on the company's attitude toward and usage of tax planning in its overall business strategy. The requirements will most likely have a meaningful effect on the extent to which companies can lower their effective tax rates as part of a general profit-creating strategy. Additionally, the IRS will be provided with information that can be used in guiding decisions including whether to audit a corporate taxpayer and which positions, if any, to challenge. Though the end result of the requirements for investors is potentially a more consistent and relevant set of income tax

information than has been previously available, it may come at the expense of increased effective tax rates. Many corporations implement tax management policies that attempt to maximize profits while minimizing reported tax expense. FIN 48 in some cases raises the reported income tax expense by requiring companies to derecognize tax benefits that would otherwise reduce tax expense. Businesses with an interest in achieving maximum tax compliance will likely face less of an impact from FIN 48 and Schedule UTP than businesses interested in minimizing tax exposure; their conservatism may actually be advantageous in reducing the likelihood of an IRS audit because the disclosure requirements allow them to credibly show a high level of confidence in their tax return. Some tax-aggressive firms may refuse to admit or disclose uncertainty at all; the responsibility of ensuring the integrity of reported uncertainty will now fall on independent auditors to uphold.

The broader notion that the government is now demanding identification of UTPs is concerning to many corporate taxpayers. Though the government must necessarily draw funding from private resources through taxation in order to provide services to a public comprised of private individuals, the United States was founded on the principle that the government ought to carry out its operations with the consent of the public. The FASB stated in the basis for its conclusions on FIN 48: "Some constituents asserted that requiring a tabular reconciliation is not appropriate because it would inappropriately provide a 'roadmap' for taxing authorities. Those constituents analogized the relationship between a taxpayer and a taxing authority to the parties in a lawsuit ... A counterparty in a lawsuit is acting in the broader public interest in regulating compliance with self-reporting income tax laws" (B64). While that statement is not incorrect, a taxing authority is privy to developing internal policies, interpretations and methods that diverge from a totally neutral reading of tax law and should not be treated as a pinnacle of objectivity and fairness. Political, societal and other pressures have the propensity to affect the IRS just as they do the FASB and public representatives.

Though it is in the public's interest for the IRS to deter noncompliance and tax evasion at both the individual and entity level, the public similarly has an interest in being able to continue to use the corporate form of organization to conduct business free of unfair and undue government intervention. "All businesses want lower taxes. But businesses

understand that their success as businesses depends in part on what the government does – on education, infrastructure, national security ... Most businesses understand that we have limited resources, that we can't raise taxes on individuals to lower business taxes and that unsustainable long-term deficits hurt growth too," offered Treasury Secretary Timothy Geithner in an interview following President Barack Obama's 2011 State of the Union address that suggested a future reduction in U.S. corporate tax rates (Wessel 2011).

In requiring corporations to report the technical results of an imprecise, broadly-conceived accounting test, the IRS is opening the door to a troubling environment where corporations with uncertain tax positions are subject to a game of courtroom 'gotcha' – the IRS can use Schedule UTP disclosures to identify tax benefits that are properly determined to be uncertain because relevant tax law is not easily or directly applicable to the event in question, and then use the very fact that the taxpayer reported such uncertainty to sway the court's opinion or pressure the taxpayer into a settlement. While it is of course entirely plausible that the IRS might indeed use enhanced disclosure solely to increase efficiency and to resolve disputes quickly, there exist at this time no clear mechanisms to protect the taxpayer from this type of underhanded government action.

Current GAAP for Recognizing UTPs in Financial Statements

FASB Interpretation No. 48 became effective as part of U.S. GAAP for fiscal years beginning after Dec. 15, 2006 and is currently incorporated into ASC 740. Under ASC 740-10, firms must evaluate each tax position to determine whether it is more likely than not to be sustained upon examination based on technical merit (740-10-55-3). Taxpayers may consider in this evaluation whether they would litigate or appeal an adverse judgment, but they are to presume that the appropriate taxing authorities have "full knowledge of all relevant information." Detection risk is ignored in the assessment of whether a position will be sustained, but widely understood administrative practices may be considered (e.g., for the sake of convenience the IRS typically ignores certain deviations from tax law such as the deduction of assets below a taxpayer-determined dollar amount) so that a company can record a related tax benefit even though the position technically would not be sustained if challenged (740-10-55-90). If any

position meets the More-Likely-Than-Not threshold, then the benefit of the position recognized in the financial statements should be as large as possible provided that amount is More-Likely-Than-Not to be realized upon settlement. Tax effects are recorded as reductions or additions to income taxes payable or as deferred tax assets or liabilities in accordance with SFAS No. 109. The recognition or non-recognition status of tax positions can change as the MLTN analysis is applied in subsequent financial reporting periods (740-10-35-3).

If a taxpayer takes positions on a tax return or expects to do so but those positions do not meet the MLTN threshold, the taxpayer must accrue appropriate interest and penalties related to those positions under ASC 740-10. ASC 740-10-25-56 requires that this interest be expensed separately of general income tax expense. In the notes to the financial statements, companies are required to report the potential changes in the status of unrecognized tax benefits, benefits related to UTPs that are taken in a tax return but not recorded because they fail the two-part FIN 48 test (740-10-50-15d). Publicly traded companies are additionally required to disclose changes in the gross amount of unrecognized tax benefits as well as the amount of unrecognized tax benefits that are realized because the statute of limitations has expired for those positions (740-10-50-15A).

Taxpayers must derecognize any and all tax benefits that do not meet the MLTN threshold. A valuation allowance is not permissible if established solely because the sustainability of the position is in question. This is significant in that the potential value of any tax benefits that fail the MLTN test are completely absent from the consolidated financial statements. Even though any potential tax benefit with a sustainability probability greater than zero theoretically holds some value to the company, many of these benefits are confined to the notes to the financial statements under ASC 740-10. If ultimately never challenged, the income effects of non-MLTN tax benefits will not be realized until all relevant statutes of limitations expire; this ordinarily takes at least three years starting from the end of the tax year in question.

Accounting Implications and Complexities

The most challenging component of FIN 48 from an accounting standpoint is probably the charge to evaluate every tax position for

sustainability upon examination. Retaining experts in tax law can be prohibitively expensive for small and medium-sized companies, and unreasonably costly for large corporations that have to contend with a multitude of potentially taxable events and a variety of operating jurisdictions, each with its own set of laws. ASC 740-10 requires tax positions to be evaluated for each “unit of account,” the most specific level at which taxpayers accumulate information to support the tax return and at which it anticipates a taxing authority will address the issue (740-10-55-85). The Codification illustrates, for example, a scenario where a research credit taken on a tax return might represent the cumulative tax effect of four separate research projects that are individually and separately substantial in scope; if documentation exists for each project, then each of the four projects would be tested using FIN 48 versus testing the aggregate research credit only. Each unit of account is tested ignoring the potential canceling or amplifying effects of interrelated tax positions. The determination of what constitutes a unit of account is subjective and dependent on relevant facts and circumstances.

Accountants also must deal with the difficult problem of trying to develop an empirical probability that a tax position will be approved upon examination. Many tax positions are widely accepted and clearly defendable using tax literature and case law, but large companies in particular are likely to find themselves wanting to take tax positions that are more problematic. It can be difficult in the latter case to quantitatively benchmark the likelihood that the IRS or another regulatory agency will accept a tax position when tax law does not yield an obvious conclusion and precedent is either nonexistent or not directly comparable to the situation at hand. FIN 48 does not prescribe a specific method of coming up with a number other than to evaluate a tax position “based on its technical merits” (FIN 48, B27). In its basis for conclusions on FIN 48, the FASB writes, “the Board does not believe that a legal tax position must be obtained to demonstrate that the more-likely-than-not recognition threshold is met. The Board believes that a tax opinion can be external evidence supporting a management assertion and that management should decide whether to obtain a tax opinion after evaluating the weight of all available evidence and the uncertainties of the applicability of the relevant statutory or case law. Other evidence, in addition to or instead of a tax opinion, supporting the assertion also could be obtained; the level of evidence that is necessary and

appropriate is a matter of judgment that depends on all available information” (FIN 48, B34).

In particular, state and local income taxes must come under substantially greater scrutiny internally as a result of FIN 48’s requirements (Kwiatek). Many states and locales utilize a complicated set of criteria to determine whether a company has nexus in that jurisdiction. Public Law 86-272 holds that states are prohibited from “imposing a tax on or measured by net income when an entity’s only connection with the state is the solicitation of orders or sales of tangible personal property.” Many states have passed statutes or regulations, though, that invoke a principle known as “economic nexus” whereby businesses can be subject to some form of income tax even if not physically present in that place. Some states have argued that intangible holding companies create an economic presence within their borders and attempt then to collect income tax from the company. Questions that states might ask in order to determine whether or not a company has nexus range from, “Does the business have an office, agency, warehouse or other business location owned or leased in the state?” to, “Does the business have any employees or representatives who use their in-state home to receive business callers?” (State Tax Nexus). CPAs with Financial Executives International observe: “A business can face multiple overlapping tax-collecting jurisdictions – countries, states, counties, cities and special districts – and tangled rules in the various tax authorities based on the type of business, location or where the products and services will ultimately be used” (Yrjanson, Tuthill 2010).

FIN 48 requires firms to identify and potentially record a liability for taxes owed to a state in which it has never before filed a return – the statute of limitations never begins to run if a return is never filed. The interpretation does allow firms to consider widely accepted administrative practice where nexus is concerned since several states have implemented tax amnesty or voluntary disclosure programs in which the “lookback” period is reduced if firms initiate contact with the jurisdiction to clarify whether they have nexus in that jurisdiction (FIN 48 A14-15, Kwiatek). Nevertheless, revisiting past years to determine whether additional liabilities ought to be recorded in accordance with FIN 48 for state and local income taxes can present administrative difficulty and will require CPAs to continually update their understanding of local tax laws or consult lawyers or tax professionals with the necessary expertise.

FIN 48 creates additional complications related to auditor independence. Accounting firms that provide both auditing and tax planning advisement services may find their perceived and individual independence compromised. It is unlikely that a CPA can be properly objective in analyzing the sustainability of a tax position that he or she advocated for the purpose of filing a tax return (Stromsem). Smaller business in particular may be forced to retain a tax professional in addition to a CPA firm as a result of FIN 48.

Achieving Consistency Using FIN 48 versus SFAS No. 5 and SFAS No. 109

According to the FASB, “Diverse accounting practices had developed with respect to the recognition and measurement of current and deferred tax assets and liabilities in financial statements [as of July 2005 when the FASB first issued an exposure draft on uncertain tax positions]. That diversity resulted from the inconsistency in the criteria used to recognize, derecognize, and measure the economic benefits associated with tax positions” (FIN 48, B2). Statement of Financial Accounting Concepts No. 2, one of the documents that guides the FASB in issuing accounting pronouncements and interpretations, advises, “Information about a particular enterprise gains greatly in usefulness if it can be compared with similar information about other enterprises ... Comparability between enterprises and consistency in the application of methods over time increases the informational value of comparisons of relative economic opportunities or performance” (SFAC 2, 3). The stated primary goal of FIN 48, then, was to achieve comparability in accounting for UTPs among various enterprises. To this end, FIN 48 is successful in that it establishes a consistent methodology that had not existed previously.

From 1992 until the effective date of FIN 48 in 2006, the prevailing guidance for recording and disclosing income tax liability had been SFAS No. 109, Accounting for Income Taxes. SFAS No. 109 superseded 1987’s FASB Statement No. 96, itself a replacement of Accounting Principles Board Opinion No. 11, issued in 1967. SFAS No. 109 made the significant change of requiring companies to record temporary book-tax differences as assets and liabilities in the balance sheet rather than as deferred charges in the income statement. SFAS No. 109 provided no guidance on how firms should record UTPs, so most companies deferred to SFAS No. 5, a generic statement on contingencies that

requires firms to accrue material losses that are “probable” and “reasonably estimated” and to disclose those that are unlikely or not estimable. With regard to UTPs, SFAS No. 5 was not always useful to financial statement users because firms enjoyed considerable flexibility when deciding when and how to disclose and accrue tax liabilities contingent upon possible audit (Gleason and Mills, 318). Investors typically have no access to IRS audit information or progress, so under the SFAS No. 5 approach of accounting for UTPs they were afforded no meaningful information related to the magnitude of any taxpayer-IRS disputes or the likelihood of favorable or unfavorable resolution of those disputes. Disagreements between taxpayers and the IRS can be ongoing until the completion of the audit and related settlements, often months or years after the end of the tax year in question.

Gleason and Mills found through interviews that the “primary consideration” in deciding whether or not to disclose an expected loss related to a tax position pre-FIN 48 was the size of the anticipated loss after settlement. Other qualities of an expected loss that led to disclosure were extremely large IRS claims, public awareness of a claim through the press and industry factors. A popular benchmark for materiality among large, frequently audited taxpaying firms was five percent of “normal income,” the greater of actual income or five percent of assets (Gleason and Mills 319). The study concluded that though firms tended to disclose the largest contingent tax liabilities voluntarily, they often under-disclosed claims that users and regulators would reasonably consider to be material. It is worth noting that Gleason and Mills ignored for the purposes of their study all unaudited returns in part “because firms may consider claims on unaudited returns to be too remote to require disclosure,” a supposition that FIN 48 markedly affected.

Another major accounting problem that FIN 48 attempted to rectify was the use of so-called “tax cushions” to manipulate earnings. “Prior to the issuance of this Interpretation,” the FASB wrote, “tax positions were sometimes recognized in the financial statements on an as-filed or to-be-filed tax basis, such that current or deferred tax assets and liabilities were immediately recognized when the related tax position was taken (or expected to be taken). In some cases, the ultimate realizability of any current or deferred tax benefit was evaluated and a valuation allowance was recorded” (FIN 48 B4). The term “tax cushion” ordinarily refers to the difference between U.S. current income tax

expense reported in the financial statements and total taxes due as reported on the corresponding income tax return (Gleason and Mills 323).

Essentially, “tax cushions” allowed a company to record for financial purposes a tax liability of, say, \$100 while only claiming \$60 of liability on the corresponding tax return, ignoring temporary differences, on the basis that the IRS might in the future disagree with the company’s treatment of a position and require payment of \$100 instead of \$60. The \$60 would be expensed in the same fiscal year as the tax return was filed – again, disregarding any temporary differences for simplicity – and the \$40 would remain on the books as a financial liability. The firm could then in a later year release the \$40 presuming that the position would be successful and record income of \$40 for that subsequent year. The tax cushion fell into the category of what Arthur Levitt termed “cookie jar reserves” in describing prominent methods of earnings management through accounting techniques (Levitt 1998). Corporate income tax returns are not publicly accessible, and so investors had no way of determining the extent to which the release of tax reserves was affecting earnings. Research suggests that some firms that might otherwise miss earning targets manage tax expense in order to meet or exceed expectations (Dhaliwal, et. al. 2004 445).

Asset/Liability Approach to Deferred Taxes

The prevailing guidance for recording the income tax effects of an event is Statement of Financial Accounting Standards No. 109, issued in February 1992. Under SFAS No. 109, tax consequences of events where no temporary difference occurs are recorded as tax expense – or a current tax asset if a refund is expected – in the same period as the taxable event. When the tax basis of an asset or liability becomes different from its reported amount in the financial statements because of a temporary difference, a deferred tax asset or liability is recorded in the balance sheet. “A deferred tax liability or asset represents the increase or decrease in taxes payable or refundable in future years as a result of temporary differences and carryforwards at the end of the current year” (SFAS 109, Summary).

The predecessor to SFAS No. 109, Accounting Principles Bulletin 11, required companies to record deferred charges in the income statement instead of in the balance sheet. The “deferred tax” account that appeared in the income statement

under APB 11 was the product of a relatively simple calculation representing the difference between the tax effects of all of the transactions undertaken in a given year and the tax owed on the corresponding tax return. No assets or liabilities were recognized, and some users treated the deferred entries as mere bookkeeping entries since no amounts were listed as being receivable or payable. APB 11 included no requirement to establish a valuation allowance for uncertain tax positions, and the effects of temporary differences were accounted for at the tax rate in effect at the time of the difference, even if the rates were expected to be different at the time of reversal. A scenario in which subscription to APB 11 would have yielded a different tax expense than subscription to SFAS 109 is illustrated in Appendix C.

Though SFAS No. 109 does require companies to record a valuation allowance, the allowance is created to capture not the sustainability of tax positions upon audit but the potential for deferred tax assets or liabilities to be unrealizable based on future circumstances. For example, a deferred tax asset related to a carry-forward loss might warrant a valuation allowance if the ability of the company to generate taxable income to be offset by the loss carried forward is not definite.

The Financial Accounting Standards Board defines assets as “probable future economic benefits obtained or controlled by a particular entity as a result of past transactions or events” (SFAC 6, 25). The first characteristic of an asset, per Statement of Financial Accounting Concepts No. 6, is that it “embodies a probable future benefit that involves a capacity, singly or in combination with other assets, to contribute directly or indirectly to future net cash inflows” (SFAC 6, 26). Using this definition, FIN 48 properly holds that a claimed tax benefit should not be recognized in a company’s financial statements if it is not probable that the benefit will be realized. Instead, a liability should be accrued representing the probable taxes payable in the future as well as any associated penalties and interest.

If the accounting treatment of the tax consequence of an event is to fairly reflect the underlying reality, though, the analysis prescribed by FIN 48 is improper as it tests not the holistic probability that a tax position will be sustained but rather the theoretical probability that a tax position would be sustained provided regulatory agencies had the same information as the taxpayer. In its

conclusions supporting FIN 48, the Board writes, “at this time, it is preferable to separately evaluate tax positions for recognition against a recognition threshold and to provide separate measurement guidance for positions that qualify for recognition” versus using an expected-outcome measurement (B26). Disregarding the effects of disclosure on the decisions of regulators, FIN 48 requires firms to record liabilities that are not actually expected.

“Implementation of the Proposed Interpretation will have the effect of consistently overstating the tax accruals for uncertain tax positions in direct conflict with the conceptual framework of the Board,” wrote one chief tax officer of a major company in a comment letter to the FASB (Stecker 2005, 4). While it is certainly possible that FIN 48 could indeed result in an overstatement of liabilities related to UTPs, such is not a necessary outcome (Mills, et. al. 2009, 26). For example, if a firm determines that the possibility of audit is 100 percent and that the likelihood that an all-or-nothing claimed tax benefit will be sustained is 60 percent, the firm will record a FIN 48 liability of zero even though there is a 40 percent chance that the full amount of the claimed benefit will be remitted to a collecting agency in the future. In such a scenario, FIN 48 actually understates the theoretical liability. When the probability of audit is greater than zero, a FIN 48 analysis may underestimate, overstate or correctly state the tax liability depending on the difference between the recognized and claimed benefit, the probability of audit and the probability that the recognized and unrecognized portions of the claimed tax position will be sustained on audit.

Though FIN 48 does require firms to disclose information that may assist the IRS in gauging the extent to which the firm has claimed tax benefits that are not likely to be sustained upon audit, the FIN 48 liability in aggregated form is not especially informative to any one government (Mills, et. al. 2009 27). While FIN 48 requires firms to test each tax position individually, they are not required to actually disclose reserves associated with any position and so the FIN 48 requirements alone could perhaps signal a government audit but probably would not spur the inspection of specific positions.

Where certain large transactions are concerned, it is possible that a reported FIN 48 liability could represent the effect of one large event. For example, companies sometimes engage in what is known as a “tax-free spin-off” when they divest a portion of their business by establishing a

brand new company with new management. Because the business is not sold, the parent company does not realize a taxable capital gain on the sale. If the tax-free status of such a transaction is questionable, the potential capital gain might wholly or largely comprise a FIN 48 liability for the parent company and in such a situation FIN 48 could by itself serve as a “roadmap” for the IRS. It is likely, though, that an IRS might be triggered regardless of FIN 48 simply because the major transaction took place at all. Additionally, FIN 48 disclosure requirements do not require a breakdown of the taxes owed to various local, state, national and international regulatory bodies.

Where FIN 48 draws the line in terms of disclosure requirements, however, Schedule UTP continues. Because Schedule UTP does require taxpayers to identify and describe each UTP individually, it is possible that its existence increases detection risk substantially. If this is true, the technical merits of a position as well as administrative practice will indeed be the only factors governing whether or not a claimed but uncertain tax benefit is realized and FIN 48 will be effective in depicting the actual tax benefits and liabilities likely to be ultimately realized. The extent to which the IRS pursues various UTPs using the information provided by taxpayers of various size – for tax year 2010, only firms with assets of \$100 million or greater will complete the form – is not yet clear. It is important to note that in the final version of Schedule UTP, the IRS will not be provided with a dollar amount of the potential reversal, but will be alerted to UTPs that represent an amount greater than 10 percent of the entire sum of reported UTPs. The IRS, then, will not have the fullest set of data necessary to perform an exact cost-benefit analysis and operate most efficiently.

Usefulness of Tax Disclosures to Financial Statement Users

At this time, it is generally impossible for a third-party user of a set of GAAP financial statements to discern with confidence the cash amount of taxes paid by a corporation to a given government for a given tax year (McGill, Outslay 2002 1131, 1136). Though the total amount of cash payments attributable to owed taxes is identified in the statement of cash flows, that number represents tax payments to multiple jurisdictions related to multiple years. U.S. tax returns are not made available to the public, so users of financial statements are forced to try to deduce an estimate

of taxes paid for a certain tax year using the cash outflow for income taxes number in the statement of cash flows in conjunction with reported income tax expense and any related footnotes. Because neither amount will necessarily approximate the cash amount of taxes paid to the IRS or any other regulatory agency, the public may jump to incorrect conclusions. In a Nov. 30, 2010 speech on the floor of the United States Senate, Sen. Bernie Sanders (I-Vt.) said, “Last year, ExxonMobil made \$19 billion in profit. Guess what. They paid zero in taxes. They got a \$156 million refund from the IRS.” The company’s 2009 10-K does indeed show that the U.S. portion of income tax expense for fiscal year 2009 was negative \$156 million, but this number represents the finalization of tax bills from previous years as well as current and deferred tax expense for 2009. *Forbes* reported that according to an Exxon spokesperson, the company will owe a “substantial 2009 tax liability” once the final tax bill is figured but declined to offer a number (Helman 2010). It is also worth noting that ExxonMobil booked \$15,165 million of non-U.S. income tax for 2009. It is not uncommon for large, international companies to siphon taxable income to countries with lower tax rates through the use of transfer pricing and complex business structures. Search engine giant Google, for example, paid an overseas tax rate of just 2.4 percent in 2009; its transfer pricing arrangements were approved by the IRS in 2006 (Drucker 2010).

Confidentiality of Corporate Tax Returns

Since 1894 (Bernasek 2010), the federal government has been prohibited from disclosing any portion of an individual or corporate tax return with limited and carefully defined exceptions. Until 1976, income tax returns were classified as public records; though not readily accessible to the public, they were permitted to be inspected by certain government agencies and parties with a material interest in the return subject to Treasury regulations and the president’s executive order (Lenter, et. al. 2003 813). The Tax Reform Act of 1976, a response to allegations that the Nixon administration had improperly used tax return information against its political opponents, eliminated executive branch control over tax return disclosure and reclassified tax returns to confidential status (Lenter, et. al. 2003 813). Today, there are few circumstances under which the IRS may release information from a tax return without the express consent of the taxpayer. The major

circumstances are: when a state tax official or state or local law enforcement agency requires such information “for the administration of State tax laws for the purpose of, and only to the extent necessary in, the administration of such laws” (U.S. Code §6103 d-1), when a shareholder owning more than one percent of a corporation requests the information (6103 e-1-D-iii), when information is requested by the chair of a Congressional committee or the President or an appointee thereof (6103 f,g), when certain federal officers or employees require information for tax administration (6103 h) and when the information is to be used in a criminal or national security investigation at the federal level (6103 i). Under none of these circumstances is the designated recipient of the information permitted to share it; i.e. a shareholder of a corporation may not obtain the corporation’s tax return legally and then publish it or disseminate information to a third party.

Current law does not allow for prospective investors, less-than-one-percent investors or creditors to access a corporation’s tax return under normal circumstances. In 2003, Rep. Lloyd Doggett (D-TX) introduced in the House of Representatives a bill titled, Corporate Accountability Tax Gap of 2003 which would “...permit inspection of true corporate tax liability and understand the tax strategies undertaken by corporations, to discourage abusive tax sheltering activities, and to restore investor confidence in publicly traded corporations” (U.S. Congress 2003). The Act would have made public certain information including net corporate income tax, taxable income and certain book-tax differences but it has not yet been brought to a vote.

Some believe that public dissemination of tax returns would be a good idea on the basis that “sunlight is the best disinfectant” or that publicizing tax returns would increase the political pressure for tax reform. Others believe that providing information to investors that extends beyond the requirements of GAAP would improve the functionality of financial markets (Lenter, et. al. 2003 814). Because the IRS is the government agency responsible for enforcing tax law and it obviously has access to return data, it is not likely that making tax returns public would improve the ability of the government to enforce laws; §6103 of the U.S. Code already allows for federal agencies to obtain relevant data from a tax return where appropriate. While it is possible that financial experts might be able to obtain information regarding income tax expense calculations beyond

what GAAP provides, FIN 48 has already reduced the ability of corporations to manage earnings through valuation allowances to the extent that viewing the tax return would not enhance markedly the quality of information required by GAAP.

An interesting argument for making at least a portion of corporate tax returns public is that to do so would encourage tax compliance because corporate managers would be wary to practice aggressive tax avoidance with the understanding that stakeholders might frown upon such behavior as evidence of bad corporate citizenship. While there is evidence that both customers and stockholders consider a social responsibility and corporate citizenship in decisions to purchase from or invest in a company, the evidence does not suggest that the financial performance of corporations is materially affected by these perceptions or that the dutiful payment of taxes is of particular importance to these stakeholders (Lenter, et. al. 2003 820). The argument is also weakened by the consideration that public disclosure of tax returns might actually incite companies to decrease their tax bills for fear of being compared by investors to competitors with smaller tax bills.

The most sustainable argument for publicizing corporate tax returns, then, is that doing so might encourage political and public pressure for fairness in the tax system by exposing inequities among companies or industries or discrepancies between the goals and outcomes of particular tax laws. While such would likely have positive outcomes, the downsides of making tax returns public need to be considered as well.

Both the Securities and Exchange Commission and the Treasury Department responded negatively to the 2002 inquiry of Sen. Charles Grassley (R-Iowa), a letter suggesting that regulators and the public might benefit from making corporate tax returns public (Lenter, et. al. 2003 804). Lenter, et. al. suggests the following as primary reasons to retain confidentiality of corporate tax returns: corporations might withhold important but sensitive information in reporting their tax liability if that information were to be publicly available, financing and operating strategies of corporations would be openly revealed to competitors and some corporations might choose to become private or even relocate to a foreign country in order to avoid disclosing such information. The issue, then, is whether the privacy of corporations ought to be trumped by the federal government's interest in establishing a fair tax system and collecting an

appropriate level of revenue in accordance with the design of that system.

Government/Taxpayer Relationship

By its nature, the relationship between taxpayers and their government – a government comprised of and controlled by taxpayers - is complicated because while it is in the interest of taxpayers in the aggregate for the government to collect enough revenue to provide the services they have, in theory, demanded of it, each individual taxpayer is inclined to pay as little as possible. The task of the IRS, then, is to perform audits at a rate that deters tax evasion, encourages maximum compliance and minimizes the cost of administration.

In fiscal year 2008, the most recent year for which comprehensive data is available, the IRS employed 20,722 enforcement officers and collected a total of \$56.4 billion through automatic underreporting techniques – for example, comparing W-2 and 1099 forms submitted from banks, investment firms and employers with the information supplied by taxpayers – and traditional examination techniques. 1.01% of individual returns were audited, with an emphasis on higher-income taxpayers; 5.57% of tax returns reporting an income of \$1 million or higher were examined, compared to 0.95% of returns reporting an income of less than \$200,000. The audit rate is low for smaller corporations – only 0.95% of corporations listing less than \$10 million in assets were audited in FY 2008, but 15.3% of corporations listing assets greater than \$10 million were examined; the rate rises to 27.4% for corporations listing assets of over \$250 million (Fiscal).

After being notified of the results of an examination, corporate and individual taxpayers have an opportunity to substantiate their claims; the IRS may accept the tax return as filed, propose a settlement with the taxpayer or disallow certain tax treatments. The taxpayer is then afforded the opportunity to contact the Appeals office of the IRS or dispute the IRS's decision with a formal trial in a court of law.

FIN 48 and Schedule UTP stand to affect nearly every step of the audit process. While the specifics of the IRS audit selection methods are purposefully not made public so as to not tip off potential tax evaders on what constitutes a red flag, the IRS has been straightforward in its message that agents will indeed use Schedule UTP to help guide decisions regarding which corporations to audit and

which tax positions in particular to examine in that audit. In April 2010, IRS Commissioner Doug Shulman said that taxpayers and tax authorities both desire a balanced tax administration system that provides “an efficient use of government and taxpayer resources by focusing on the issues and taxpayers that pose the greatest risk of tax noncompliance” (Shulman 2010). In that same address, Shulman stated, “An important thing for you to realize is that a major goal of this proposal [an early draft of Schedule UTP that required taxpayers to list a maximum tax adjustment – that requirement was removed from the final instructions] is to use the schedule for audit selection, not just as information in audits.”

Taxpayers are also concerned that Schedule UTP implicitly provides the IRS with privileged information. In preparing their financial statements in accordance with GAAP, corporations must frequently consult the expertise and counsel of a tax attorney, an accounting firm, or both so that it can support the MLTN status of claimed tax benefits recorded in the financial statements to its financial statement auditor. Schedule UTP requires taxpayers to identify not only benefits that fail the MLTN test on their technical merits but also benefits that would fail the MLTN were it not for the firm’s intent to litigate an adversarial opinion of the IRS. In a very real sense, then, a corporation is required to unveil to the IRS the opinions of its legal counsel – that is, the opinion that a position is not MLTN to succeed on technical merit even if it can, perhaps, succeed on legal merit – before any sort of trial begins. As one scholar articulates, “Ambiguity in the tax law is resolved by litigation – a system of adversaries. For effective common law to develop, courts must hear ‘zealous advocacy’ from both sides in every case. Tax law is no exception.” (Jones 799).

Empirical research finds that many companies released a material amount of tax reserves shortly prior to the effective date of FIN 48 (Blouin, et. al. 2006, 808). The research supports the notion that many companies desire to minimize their FIN 48 liability. A desire to reduce the kind of IRS scrutiny that might weaken a corporation’s negotiating positions is a probable contributing factor to this phenomenon; the authors also consider a desire of companies to record the release of any lingering excessive tax reserves as an increase in income statement earnings versus as a cumulative change in accounting principle adjustment to retained earnings that would have no effect on current-year income.

It should be noted that FIN 48 and Schedule UTP do not necessarily yield a disadvantage for every corporate taxpayer in every circumstance. On the contrary, the enhanced disclosure requirements for the first time allow a taxpayer to credibly indicate to the IRS that all or most of the positions taken in the tax return are MLTN to be sustained. “FIN 48 re-enforced my belief that taking risky tax positions is not in a company’s best interest,” wrote one tax professional wishing to remain anonymous. “I realize there are grey areas and differences of opinion can exist, but I’ve never lost an issue I’ve researched and built a solid supporting position for ... My comfort level for taking a position is closer to a ‘should’ position and thus a FIN 48 reserve would be zero.” The 2009 paper by Mills, et. al. finds that “taxpayers with strong positions have higher expected payoffs post-FIN 48 than they did pre-FIN 48” (29) because the government has less incentive to audit a company showing little or no uncertainty under FIN 48 compared to a company that provides no credible disclosure of position sustainability at all, and certainty compared to a company showing a high level of uncertainty.

Observed Attitudes of the IRS

In 1984, the U.S. Supreme Court ruled that the IRS may obtain tax accrual workpapers under 26 U.S.C. §7602, which authorizes the Secretary of the Treasury to summon and “examine any books, papers, records, or other data which may be relevant or material” to a particular tax inquiry (United States v. Arthur Young & Co., 813). The Court found that, “It is the responsibility of the IRS to determine whether the corporate taxpayer, in completing its return, has stretched a particular tax concept beyond what is allowed. Records that illuminate any aspect of the return – such as the tax accrual workpapers at issue in this case – are therefore highly relevant to legitimate IRS inquiry” (815). In that same case, the Court considered whether some sort of accountant-client privilege ought to be enacted, but found that because independent auditors, unlike lawyers, ultimately bear responsibility to the public over the client, “To insulate from disclosure a certified public accountant’s interpretation of the client’s financial statements would be to ignore the significance of the accountant’s role as a disinterested analyst charged with public obligations” (818).

Since the Supreme Court ruling, the IRS has affirmed a policy of restraint with regard to tax accrual workpapers. “Audit or tax accrual

workpapers should normally be sought only when such factual data cannot be obtained from the taxpayer's records or from available third parties, and then only as a collateral source for factual data," the IRS writes in its Internal Revenue Manual (4.10.20.3.1). "Audit or tax accrual workpapers should be requested with discretion and not as a matter of standard examining procedure." This policy was first asserted in 1984, but was adjusted in 2002 to exclude "listed transactions" from restraint as a response to the growing preponderance of high-profile corporate tax shelters. After FIN 48 became effective in December 2006, there was much concern within the accounting profession that the IRS would be tempted to utilize FIN 48 workpapers and renege on its self-imposed policy of restraint. In May 2007, the IRS published a "FIN 48 Implications LB&I Field Examiners' Guide" which noted: "The disclosures required under FIN 48 should give the Service a somewhat better view of a taxpayer's uncertain tax positions; however, the disclosures still do not have the specificity that would allow a perfect view of the issues and amounts at risk." The guide continues, "Even with the lack of specificity, tax footnotes included in financial statements, including FIN 48 disclosures, should be carefully reviewed and analyzed as part of the audit planning process ... Revenue Agents should not be reluctant to pursue matters mentioned in FIN 48 disclosures, but should be mindful of our policy of restraint on Tax Accrual Workpapers and not cross over the boundaries contained there." The guide also notes that some taxpayers desire expedient resolution of uncertain tax issues since a large FIN 48 liability can have a negative impact on the financial statements.

In 2009, the issue of work product privilege was applied to FIN 48 in USA v. Textron. Textron is a publicly traded company audited by Ernst & Young; in 2003, the IRS audited Textron's 2001 tax return and issued an administrative summons to obtain tax accrual work papers under §7602 because it was suspected by the IRS that Textron had engaged in sale-in, lease-out transactions as well as other "listed transactions." The U.S. Court of Appeals for the First Circuit ruled that Textron's work papers are not protected under the work product privilege because tax accrual work papers are prepared primarily for the sake of calculating items to be included in financial statements and not in anticipation of a court trial (24). The Supreme Court decided not to review the appeal, letting the ruling of the appeals court stand.

The 2009 Textron case differs from the 1984 Arthur Young case in that Textron's work papers were produced by Textron's tax department with consultation from outside legal counsel; in the Arthur Young case, the work papers were prepared by the auditor and so the issue of work product privilege was never considered. The driving contention in the Textron case is not the relevance of the tax accrual work papers to the IRS audit but whether or not the work papers can be considered documentation that is prepared because of the prospect of litigation. Essentially, the First Circuit court's ruling holds that the work required to document a FIN 48 analysis for the sake of proving proper financial statement to auditors can be requested and summoned by the IRS in order to determine whether or not a tax benefit holds merit. Because managers of publicly traded companies need to support the assertions contained in financial statements in order to obtain an unqualified opinion from an auditor, they are left no choice but to prepare documentation of uncertain tax positions that incorporates the opinions of legal counsel knowing that such documentation could ultimately be used against them by the IRS.

In response to these concerns, the IRS in October 2010 issued Announcement 2010-76, which expanded the Service's existing policy of restraint; the announcement affirms the ability of agents to request documentation of legal support for positions, but permits taxpayers to withhold a document if it "is otherwise privileged under the attorney-client privilege, the tax advice privilege in §7525 of the Code, or the work product doctrine and the document was provided to an independent auditor as part of an audit of the taxpayer's financial statements." That policy became effective Sept. 24, 2010. Once again, the policy of restraint does not apply if unusual circumstances exist or if the taxpayer has claimed the benefits of one or more "listed transactions." The IRS also permits taxpayers to redact information from requested tax reconciliation work papers if that information contains specific calculations and reserve amounts related to Schedule UTP.

On March 23, 2011, the IRS issued a "Frequently Asked Questions on Schedule UTP" on its website; the document clarified that the policy of restraint outlined in Announcement 2010-76 applies to documents requested by Appeals as well as documents requested by Counsel after a Tax Court petition has been filed by a taxpayer.

Initial Schedule UTP Proposal and Feedback

On April 19, 2010, the IRS released a draft of Schedule UTP that would have, if fully enacted without revisions, required corporations listing assets of \$10 million or greater to disclose UTPs to the IRS beginning in tax year 2010. The initial draft required filers to complete the same analysis as is required in the final version, but also required corporations to calculate and report a Maximum Tax Adjustment for each UTP – the maximum amount in dollars that might be retained by the IRS should the IRS prevail in challenging the position.

After an open comment period, the IRS made adjustments to the draft that eliminated the Maximum Tax Adjustment requirement, allowed corporations to exclude UTPs that were not reserved under FIN 48 because of administrative practice and scaled back the filing requirements so that only those companies listing \$100 million or greater in assets would be subject to the requirements immediately. Corporations listing \$50 million in assets will fall under the requirements beginning in tax year 2012, and corporations listing \$10 million will begin filing Schedule UTP in tax year 2014. A final major adjustment in the final draft is the elimination of the requirement for corporations to report to the IRS the rationale for taking an uncertain position. Critics had argued that such a requirement would force corporations to reveal privileged information to the IRS. The final version requires corporations to explain the nature of the uncertainty for each UTP but does not ask them to justify their decision to take the position.

The upsides of Schedule UTP from a public perspective are not trivial. The schedule will allow the IRS to audit more efficiently by concentrating its resources on resolving those issues that are most likely to exploit inconsistencies and ambiguities in the tax code. The manager of accounting policy and financial research for FEI writes that a review of reported UTPs will allow the IRS Large Business & International Division to potentially undergo the process of “publishing guidance necessary to eliminate uncertainty wherever possible, as well as identifying areas for possible legislative changes” (Wei 2010). Both of these proposed uses of Schedule UTP would be legitimate and welcome. There is concern, though, that another use of the schedule involves the executive branch equivalent of judicial activism – proactively challenging tax positions in court or in negotiation in such a way as to maximize present and future collections instead

of striving for consistent, objective interpretation of the law.

Interactions Between Schedule UTP and Restraint Policy

IRS Commissioner Doug Shulman stated to the American Bar Association that the IRS will not seek documents that would otherwise be privileged, nor will it procure drafts of issue descriptions or information regarding the ranking of issues used in preparation of Schedule UTP (Shulman 2010). The stated goal of the IRS is to create certainty for the government and taxpayers alike sooner and to operate efficiently and effectively. In a September directive, the Deputy Commissioner for Services and Enforcement of the IRS instructed personnel to “approach UTPs on audit keeping in mind their responsibility to apply the law as it currently exist, not how we would like it to be. We must do this without bias in favor of the government or the taxpayer” (Miller 2010). Though these words soothe the fear that the government will use Schedule UTP to exploit ambiguities in tax law, the structural system created by the policy relies heavily on a fallible promise. Once a corporation has acknowledged that a tax position is uncertain, the IRS has full reign to use that information to sway a court.

The LB&I Subgroup of the IRS Advisory Council reported in November 2010 that it disagrees with the issuance of Schedule UTP and finds “significant challenges to overcome to make sure that examination teams utilize the information contained on this schedule in a reasonable manner” (IRS Advisory Council). There is also legitimate concern about the consequences of Schedule UTP contents being released to foreign governments. The release of tax information between countries can trigger heavy scrutiny of transfer pricing arrangements that affect the taxability of income in various international jurisdictions, and jurisdictions outside of the U.S. might use the information aggressively to the disadvantage of corporations. The IRS has responded to this concern by noting that only under very limited conditions would release of Schedule UTP data to other jurisdictions be considered.

Role of Auditors in Government-Taxpayer Relationship

The largest U.S. corporations almost exclusively employ four global accounting firms:

Deloitte & Touch, Ernst & Young, KPMG and PriceWaterhouseCoopers. These firms have collectively become known as the “Big Four” because they collectively provide advisory, tax and auditing services to a vast majority of publicly traded and large private companies. The SEC does not require listed companies to choose these firms as auditors, but the Big Four have established globalized capabilities and competencies that regional firms cannot match. The Sarbanes-Oxley Act of 2002 forbids any accounting firm from simultaneously auditing and providing certain non-assurance to publicly-traded corporations. Many publicly-traded corporations therefore utilize two Big Four firms at once, each providing a different set of services. Considering that two international companies engaged in a merger or acquisition will likely each wish to retain one accounting firm to perform auditing services and one to perform due diligence services, a minimum of four global accounting firms must exist to provide support services to the two companies as they join together.

Accounting firms play a major role in determining the extent to which the IRS can successfully identify UTPs by certifying that the company’s FIN 48 analysis has been sound. From an advisement perspective, accounting firms providing tax services can sell tax shelter packages or instruct tax avoidance strategies and provide legal and technical support for uncertain positions that can be used to justify categorizing a tax position as MLTN to be sustained upon audit. From an auditing stance, the firms bear the responsibility of certifying that a company’s financial statements comply with GAAP and, by extension, ASC 740. Because the MLTN test involves a subjective judgment of the empirical justifiability of a tax position, the opinions of the tax advisor and auditor directly dictate whether or not a company will be required to report a given tax position to the IRS on Schedule UTP and subject that position to special scrutiny. The value of Schedule UTP, then, depends completely on how accounting firms interpret and enforce ASC 740 and the MLTN test specifically.

Because the Big Four exercise what amounts to an oligopoly over providing accounting and auditing services to the largest U.S. corporations, government agencies have a limited ability to sanction firms that violate policies. The 2003 KPMG tax shelter investigation illustrates this perfectly. Though the Senate found that KPMG had indeed knowingly sold tax shelters to clients, the government agreed to dismiss criminal prosecution of the firm so long as it paid \$456 million in fines,

restitution and penalties and complied with other conditions including the termination of two tax practice areas (IRS Notice 2005-83). Following the 2002 collapse of Arthur Andersen, the risk of another global accounting firm shutting its doors posed too great a risk to the economy and so the government instead held KPMG to an impossible ultimatum. The Department of Justice’s charges against 13 individual KPMG employees were later thrown out after a judge ruled that the U.S. improperly pressured KPMG to not pay the legal fees of its former employees. The case illustrates the tension between the desire of the government to influence the practices of accounting firms and the competing desire to have effective competition among large accounting firms. Though accountants and the FASB are in theory independent of the government and of clients, both are subject to political pressures. Accounting policies and disclosure requirements can either undermine or aid the enforcement techniques of governmental agencies and particularly the IRS, as illustrated by the Service’s use of FIN 48 requirements to justify the issuance of Schedule UTP. Accounting practices and standards do not exist in a vacuum; they need to be useable and relevant to a system of stakeholders with different and sometimes conflicting needs.

The fact that the collapse of any one of the Big Four firms would pose a great risk to the global and domestic economy holds foreboding connotations for a public that depends on their auditing services. Though regulatory agencies like the IRS and SEC have the ability to sanction or press charges against corporations, they rely heavily on the notion that corporate financial statements have undergone reasonable verification by a Big Four firm.

Related Topics in Accounting

Accounting for income taxes is distinguished from accounting for most other types of transactions because income taxes are calculated using non-GAAP conventions. In discussing the issues associated with providing information to third-party users based on accounting estimates when the third-party user’s reaction to and use of that information may well affect the extent to which the original estimate becomes a certainty, though, it is useful to consider the accounting treatment of contingent liabilities associated with litigation, the environment, and multiemployer plans. Additionally, consideration must be given to the fact that the U.S. will likely adopt International

Financial Reporting Standards to replace GAAP within this decade. Should the change come to pass, differences in accounting for uncertain income tax positions will affect the information collection techniques of the IRS.

Litigation Exposure

SFAS No. 5 provides guidelines that are intended to apply directly to the accounting for pending or threatened litigation. As mentioned earlier, the standard requires companies to accrue a loss if and only if the incurrence of a liability or impairment of an asset is probable and the amount of the loss can be reasonably estimated. The standard defines probable as: "The future event or events are likely to occur" (SFAS 5, 3a). If a loss is not accrued but there is a reasonable probability that a loss has occurred, disclosure must be made indicating the nature of the contingency and the range of the loss, if an estimate can be made (10). Contingent gains are never recorded in a company's financial statements, though careful disclosure must be made in the case that a gain is probable or reasonably possible (17a, b).

The FASB clarifies in SFAS No. 5 that, "The filing of a suit or formal assertion of a claim or assessment does not automatically indicate that accrual of a loss may be appropriate. The degree of an unfavorable outcome must be assessed" (SFAS 5, 37). The statement appropriately suggests that accrual of a contingency should be based on an independent, technical analysis of the situation and not on the formal actions of an antagonist party. Under this thinking, corporations should not write off tax benefits that have been taken under the belief that they are consistent with prevailing tax law simply because the IRS or another regulatory agency has challenged the validity of those benefits.

The current guidelines for recording UTPs differs from the guidelines for recording the potential effect of future lawsuits in that tax benefits must be scrutinized even if management does not believe that the benefits will be challenged. SFAS No. 5 states: "With respect to unasserted claims and assessments, an enterprise must determine the degree of probability that a suit may be filed or a claim or assessment may be asserted and the possibility of an unfavorable outcome ... If the judgment is that assertion is not probable, no accrual or disclosure would be required" (38). Corporations need not disclose or accrue losses related to litigation that is unlikely to

ever be threatened because the potential plaintiffs are unaware that they have a valid complaint or choose not to press charges. Indeed, it would be nonsensical for a company to disclose and accrue a loss related to a lawsuit that would not even be brought were the company not openly admitting to having wronged another party. In such a case, the accounting framework would actually lead to the creation of a liability rather than being the mechanism that records it.

Where taxes are concerned, though, GAAP is requiring companies to do just that – create a liability that is nonexistent for practical purposes because of a theoretical application of accounting standards. Accruing or disclosing losses or foregone benefits related to UTPs would be appropriate only if the likelihood of those positions being audited is 100% independently of the loss disclosure.

While disclosures related to ongoing litigation do have the potential to give an opposing party's counsel insight into whether the company considers a loss reasonably possible or even probable, the judgment surrounding those disclosures is based on circumstances as they exist at the time the financial statements are compiled, not as they might exist after the statements are released. ASC 740 in conjunction with Schedule UTP creates an iterative effect whereby the accounting treatment affects the event being accounted for – a very troubling side effect considering the information provided by financial reporting is intended to be historical. The FASB writes: "To the extent that financial reporting provides information that helps identify relatively efficient and inefficient users of resources, aids in assessing relative returns and risks of investment opportunities, or otherwise assists in promoting efficient functioning of capital and other markets, it helps to create a favorable environment for capital formation decisions. However, investors, creditors, and others make those decisions, and it is not a function of financial reporting to try to determine or influence the outcomes of those decisions" (SFAC 1, 33). The FASB addresses taxing authorities specifically: "Although both taxing authorities and rate-making bodies often use the information in financial statements for their purposes, both also have statutory authority to require the specific information they need to fulfill their functions and do not need to rely on information provided to other groups" (26).

Environmental Liabilities and Asset Retirement Obligations

FASB Interpretation No. 47, like FIN 48, was issued in response to companies' adoption of diverse accounting practices related to AROs. FIN 47 clarifies that companies must record the fair value of conditional AROs as soon as incurred if the fair value can be reasonably estimated. In some cases, disclosure provided by the company is supplemented with and even dictated by data that is publicly available through the Superfund Program. Superfund refers to the 1980 Comprehensive Environmental Response, Compensation, and Liability Act that gave the Environmental Protection Agency the authority to identify parties responsible for the contamination of U.S. sites and compel cleanup efforts.

An interesting facet of the accounting guidelines related to AROs involves the characterization of a liability as broader than a contractual obligation to sacrifice a resource. Statement of Financial Accounting Standards No. 143 states that its guidelines apply to "legal obligations ... a legal obligation is an obligation that a party is required to settle as a result of an existing or enacted law, statute, ordinance, or written or oral contract or by legal construction of a contract under the doctrine of promissory estoppel" (2). SFAS 143 applies promissory estoppel as defined by Black's Law Dictionary: "The principle that a promise made without consideration may nonetheless be enforced to prevent injustice if the promisor should have reasonably expected the promisee to rely on the promise and if the promisee did actually rely on the promise to his or her detriment" (SFAS 143, A2c). The implementation guidance notes correspondingly that "a legal obligation may exist even though no party has taken any formal action" and requires entities to base their evaluation of whether an obligation exists on current laws and not on forecasts of future laws or changed interpretations of existing laws (SFAS 143, A3). The FASB cites the example of a CEO who makes public comments stating that his company will clean up an abandoned building in future years as potentially constituting a recordable liability under promissory estoppel. In its basis for conclusions, the FASB writes: "Once an entity determines that a duty or responsibility exists, it will then need to assess whether an obligating event has occurred that leaves it little or no discretion to avoid the transfer or use of assets" (SFAS 143, B29).

Again, while the public would like for all companies to voluntarily identify and clean up any environmental damage they cause, it would not make sense for companies to report to investors and creditors financial liabilities that have a remote probability of ever materializing. It is possible that laws becoming effective in the future will create environmental liabilities for certain companies, but to record an obligation related to a potential-but-not-certain legal requirement would not reflect the underlying realities in which the company operates and SFAS 143 appropriately requires that judgments be made in accordance with existing interpretations of existing laws. Financial statements should reflect the financial and economic position of a company as it really is, not as it would be under ideal or contrived circumstances. Ideally, all customers would satisfy debts owed to a company and an allowance for bad debts would be unnecessary; in a worst-case scenario, all debts would prove uncollectible and revenues would be recorded only when cash is received. Neither scenario is the most probable, though, and the current accounting system strives to best represent the actual probability of collecting the receivables rather than defer to extreme possibilities. Moreover, the probability that individual accounts will be collected is typically estimated using factors like the creditworthiness of the debtor, the size of the debt, the debtor's payment history and the length of time that has elapsed since the debt arose. The obligation to satisfy the debt has no relation to whether or not the company believes the debt is collectible. Similarly, the obligation to pay taxes associated with income realized in a given transaction or event is unrelated to whether or not the company or the IRS believes that obligation is real.

Of course, the enforcement of tax law by the IRS is usually the predictor of tax liabilities, but the associated actions would not create a liability were there not a legal framework detailing and mandating taxation policies in the first place. There are two factors that predicate the recording or disclosure of a liability where legal or environmental liabilities are concerned: that the liability arises out of a valid mechanism – an on-the-books law – through which a third party including but not limited to a governmental enforcement agency can demand payment from the company, and that some third party, known or unknown to the entity issuing the statement, will be aware of that mechanism and use it to seek payment.

Participation in Multiemployer Plans

The FASB is expected to issue a final document related to enhancing the disclosure requirements about an employer's participation in a multiemployer plans in the second quarter of 2011 (FASB Current Technical Plan). Existing guidance related to accounting for multiemployer plans is represented in Codification Subtopic 715-80. The FASB writes that it decided to update 715-80 after it received "comments from various constituents on the perceived lack of transparency about an employer's participation in a multiemployer plan" (FASB Exposure Draft). The proposed amendments would require employers to provide a narrative description of "the employer's exposure to significant risks and uncertainties arising from its participation in the plan(s). That narrative description shall include the extent to which, under the terms and conditions of the plan(s), the employer can be liable to the plan(s) for other participating employer's obligations." Employers would also be required to identify "known trends in contributions, including the extent to which a surplus or deficit in the plan may affect future contributions," and other related pieces of information.

Critics of the proposed amendment argue that it will discourage employers from providing defined benefit plans to employees and instead offer defined contribution plans because the amendment will force companies to disclose liabilities that are unlikely to be realized. In particular, the amendment would require companies to disclose the effects of their potential withdrawal from a multiemployer plan even if they have no intention of withdrawing. "As stated in the FASB's Guiding Principles, to be neutral, 'information must report economic activity as faithfully as possible without coloring the image it communicates for the purpose of influencing behavior in any particular direction,'" one critic writes (Kraw). "The FASB proposals, if adopted, are by their very construct biased against the continuation of multiemployer defined benefit plans. They inevitably will produce large amounts of inaccurate information that exaggerate pension plan liabilities, encourage short term thinking and push employers to exit." Other letters addressed the fact that information about assets and accumulated benefit obligations of defined benefit plans is difficult to collect and contingent upon decisions made by labor unions with little forewarning and thus would be dated and allegedly useless if presented in financial

statements. One critic described the proposed disclosures as "inherently capable of misinterpretation, at best, and utterly meaningless, at worst" (Potts-Dupre).

Though the issues and potential impacts of the proposed multiemployer plan disclosure amendment are different from those associated with uncertain tax positions, the debate regarding when, if ever, to report prospective liabilities that have a remote chance of actually occurring. While it is true that companies choosing to withdraw from an underfunded pension plan must typically pay large withdrawal liability assessments or "exit fees," it is unreasonable to ask companies to report those amounts if the chance of ever having to pay those amounts is remote. Many businesses are susceptible to the impact of natural disasters and war, but a quantification of those kinds of risks would be misleading to a prospective investor if the businesses were otherwise financially secure.

Convergence with International Standards

The U.S. and the FASB are in discussions with the International Accounting Standards Board to eventually converge the guidelines of GAAP with those of International Financial Reporting Standards so that there will exist a set of unified global accounting policies. The SEC has indicated that the largest U.S. corporations will not be required to adopt IFRS until 2015 or 2016 at the earliest, if at all (Tweedie), and a final decision on whether or not to officially incorporate IFRS into the U.S. domestic reporting system may come during 2011.

If convergence does occur, it may yield an entirely new set of standards related to accounting for UTPs. Should ASC 740 be altered, Schedule UTP would have to be reevaluated. It is currently based on the MLTN test, a test that could be superseded or altered by new income tax accounting requirements. Schedule UTP is now justified on the basis that companies do not incur significant costs in preparing Schedule UTP because they already analyze the sustainability of tax positions in order to comply with GAAP. If the requirement to analyze the sustainability of all tax positions ceases, Schedule UTP compliance would become more costly for companies.

International Accounting Standard No. 12, “Income Taxes,” currently advises that unresolved disputes with tax authorities be handled in accordance with IAS 37, “Provisions, Contingent Liabilities and Contingent Assets” (IAS 12, 88). IAS 37 advocates an expected outcome approach that uses the probability-weighted average of a variety of possible outcomes to calculate the amount that should be recorded as a tax benefit or liability. The standard requires separate treatment of provisions, defined as “liabilit[ies] of uncertain timing or amount” and contingent liabilities, defined as “a possible obligation that arises from past events and whose existence will be confirmed only by the occurrence or non-occurrence of one or more uncertain future events not wholly within the control of the entity” or a present obligation not recognized because resource outflow is improbable or the amount of the obligation is uncertain (IAS 37, 10). UTPs would likely be treated as contingent liabilities because the obligation to pay the tax is dependent upon the actions of regulatory agencies and courts.

IAS 37 does not require entities to recognize contingent liabilities (27), but does require disclosure “unless the possibility of an outflow of resources embodying economic benefits is remote” (28). Contingent liabilities must be continually assessed and recognized as a provision in the financial statements “if it becomes probable that an outflow of future economic benefits will be required for an item previously dealt with as a contingent liability” (30). The standard instructs: “Where the provision being measured involves a large population of items, the obligation is estimated by weighting all possible outcomes by their associated probabilities ... The provision will therefore be different depending on whether the probability of a loss of a given amount is, for example, 60 per cent or 90 per cent. Where there is a continuous range of possible outcomes, and each point in that range is as likely as any other, the mid-point of the range is used” (39). Disclosure must include a brief description of the nature of the obligation and an indication of the uncertainties about the amount or timing of outflows (85). The standard does not explicitly state the extent to which individual liabilities must be separately identified, stating, “It is necessary to consider whether the nature of the items is sufficiently similar for a single statement about them to fulfill the requirements” (87).

At a July 2005 meeting of the IASB, the Board agreed that enterprises should presume that taxing authorities will review a tax positions when evaluating whether the position is probable of being sustained and that consideration of detection risk is inappropriate (IASB Agenda Project). An exposure draft proposing to replace IAS 12 was issued in March 2009. It included the following guidance: “Uncertainty about whether the tax authorities will accept the amounts reported to them by the entity affects the amount of current tax and deferred tax. An entity shall measure current and deferred tax assets and liabilities using the probability-weighted average amount of all the possible outcomes, assuming that the tax authorities will examine the amounts reported to them and have full knowledge of all relevant information. Changes in the probability-weighted average amount of all possible outcomes shall be based on new information, not a new interpretation by the entity of previously available information” (Exposure Draft 26).

In November 2009, the IASB found that support from respondents was limited and decided that the project would not proceed in its then-current form (IASB Agenda Project). On its website, the IASB writes: “The project originally started as a convergence project with US GAAP. However, in the light of responses to an exposure draft published in 2009, the Board has narrowed the scope of the project. The Board may consider a fundamental review of the accounting for income taxes after 2011” (Work plan for IFRSs). Though the Board intends to consider uncertain tax positions separately from IAS 12 as a whole, it will not visit the issue until the revision of IAS 37 is finalized; that project will not be finished until after June 2011 according to the work plan website.

Another point of distinction to be resolved between GAAP and IFRS involves the reporting threshold. While ASC 740 disallows companies from recording any portion of a tax benefit that is less than 50% likely to be sustained, IFRS tends to utilize an expected-outcome method. Consider a tax benefit of \$100 that the company believes has a 20% chance of being completely sustained, a 40% chance of being reduced to \$60 and a 40% chance of being completely disallowed. The FIN 48 test would require that the company record a tax benefit of \$60, since the probability that the benefit will be at least \$60 is 60%: 40% plus 20%. An expected-outcome calculation would yield a tax benefit of \$48: $0.20 \times 100 + 0.40 \times 60 + 0.40 \times 0$.

If there was no chance that the \$100 benefit would be disallowed entirely but a 40% chance it would be reduced to \$50, the FIN 48 test would still yield a \$60 tax benefit but an expected-outcome approach would yield a recordable benefit of \$62: $0.20 \times 100 + 0.40 \times 60 + 0.40 \times 50$. Though an expected-outcome calculation incorporates a wider range of possible outcomes, critics argue that it requires companies to estimate the probability that remote and highly certain positions will be sustainable. The Tax Executives Institute commented on the IAS 12 exposure draft: “The absence of a recognition threshold (as proposed in the Exposure Draft) would result in potential inaccuracies by requiring companies to recognize tax benefits that are highly uncertain or even those for which no or only meager authority exists. Likewise, reserves would have to be provided for highly certain positions if there are possible outcomes yielding less than 100 percent of the tax benefits. The measurement of tax positions that are at the highly certain or highly uncertain ends of the probability spectrum should not be skewed by outcomes that are remote” (TEI Comments).

IAS 12 is silent on UTPs in its existing form, so the ultimate IFRS will likely have a significant impact on the future of Schedule UTP if the U.S. does adopt international reporting standards in the future.

Conclusions

The FASB defines a liability as a “probable future sacrifice of economic benefits arising from a present obligation.” The legal requirement to pay taxes clearly creates a liability, and stakeholders of a company need to be aware of that company’s liabilities in order to make an informed decision regarding whether or not to enter into a financial relationship with the company. What FIN 48 attempts to address, though, are those situations in which it is unclear whether or not the obligation to pay taxes exists. U.S. tax law includes a large number of provisions governing income and expense classification as well as policy-driven deductions, credits and penalties that require tremendous compliance efforts. When tax laws do not neatly correspond to the events and transactions undertaken by a given corporation, the managers of that business must interpret applicable laws and decide whether or not to report taxable income.

If corporations report to the IRS and other regulatory agencies tax benefits that are likely to be challenged and reversed, investors should be made

aware of those potential liabilities. The problematic side effect of reporting such liabilities, though, is the possibility that doing so may actually trigger scrutiny of the positions that would not have taken place absent the financial reporting. The statute of limitations for tax reporting far exceeds the time period companies are afforded to prepare financial statements, and so regulators will always be able to refer to a company’s financial statements in inspecting that company’s tax return. Though FIN 48 was originally intended to clarify reporting practices related to income taxes and has been successful in creating more consistency than existed previously, it has also created a tool through which regulators can garner a peek into the perspective of the taxpayer – an advantage not reciprocated. The requirements of FIN 48 make it far more difficult for tax-aggressive companies to underreport tax liability, but at the cost of potentially influencing, for better or worse, the economic realities that GAAP purports to reflect.

While the disclosures required by FIN 48 are theoretically flawed because detection risk is ignored, an aggregated presentation of potential tax liability is not egregiously harmful. Though some distortion occurs in estimating tax position sustainability without regard to audit probability, the distortion is likely less significant than that created by the misuse of reserves and reporting policy discrepancies that were permissible under SFAS 109.

Of greater concern from a policy standpoint is the IRS’s new Schedule UTP. Though the schedule will undoubtedly improve the U.S. government’s ability to target its audits toward likely tax evaders and thereby increase corporate collections, it holds the capacity to alter the dynamics of the taxpayer-government relationship significantly. Even if the IRS upholds its policy of restraint and chooses not to seek explicitly privileged documentations and opinions, it will now have at its disposal a tool that allows it to target those tax benefits that a company’s managers, lawyers and accountants believe are susceptible to successful challenge, even if those parties all believe that the position has standing supportable by law. While decisions regarding tax law will ultimately be made in the courts, the fact that the IRS will have undue insight into the strength of taxpayer positions means it can potentially strong-arm companies into reaching settlements or sway a judge knowing that the company’s management has exposed internal doubts on the sustainability of a position.

Of course, that the IRS merely has the ability to take maximal advantage of Schedule UTP disclosures does not mean that it will do so, just as not all corporations will play the “audit lottery” or take potentially risky tax positions simply because they have the option to do so. Checks on each party are important to maintaining the integrity of a taxation system born out of democratic processes and based on self-reported liabilities versus government-imposed assessments. The IRS – an agent of the public itself – needs to retain the ability to effectively audit individual and corporate taxpayers so that tax burdens are effected in the manner that the legislative branch of government has instructed. Similarly, taxpayers must be afforded the ability to dispute IRS decisions and have contentions resolved by a neutral third party, the judicial branch. These checks allow the need of both parties to pursue their respective interests to be balanced by a public-demanded fairness.

If the IRS is going to ask corporate taxpayers to disclose information intended for internal use, the agency should provide a similar level of disclosure regarding its interpretations of its own policies and case assessment strategies. Taxpayers should have plentiful opportunities to solicit guidance from the IRS and seek preliminary interpretations of policy without the possibility that such inquiries will color subsequent, formal assessments. In 2005, the IRS piloted the Compliance Assurance Process, a program that allowed some large business taxpayers to identify and resolve complicated tax issues prior to filing their tax return. The program allows the IRS to use audit resources more efficiently, gives the taxpayer more certainty and a reduced compliance burden and provides financial statement users with more precise and reliable tax information than is available when a company’s tax return is completely subject to challenge by regulatory agencies. The CAP, though, is currently optional and Schedule UTP is not. The IRS should continue to make guidance publicly available so that taxpayers, if they so choose, can satisfy themselves as to a tax position’s sustainability before taking the position and deciding whether or not to record it in the financial statements.

The issues associated with accounting for income taxes also need to be part of a larger, ongoing conversation about revisions to the U.S. tax code. Many cite the financial burden of complying with the current system and high marginal corporate tax rates as reasons to modify the system so that top marginal rates are lower but

fewer deductions and credits are available. A simpler tax code could reduce the incidence of uncertain tax positions altogether, a trend that would benefit the IRS, taxpayers and financial statement preparers and users.

Works Cited

- American Institute of Certified Public Accountants, “Guiding Principles of Good Tax Policy: A Framework for Evaluating Tax Proposals.” March 2001.
- Bank, Steven. “Entity Theory as Myth in the Origins of the Corporate Income Tax.” 43 Wm. & Mary L. Rev. 447 (2001), <<http://scholarship.law.wm.edu/wmlr/vol43/iss2/3>>
- Bartlett, Bruce. “The Federal Tax Debate: How Much Should Corporations Pay?” Prepared for The Heritage Foundation. Jan. 8, 1985.
- Bernasek, Anna. “Should Tax Bills Be Public Information?” *The New York Times*. Pub. Feb. 13, 2010.
- Blouin, Jennifer, Cristi Gleason, Lillian Mills and Stephanie Sikes. “Pre-Empting Disclosure? Firms’ Decisions Prior to FIN No. 48.” *The Accounting Review*. May 2010.
- Congressional Budget Office Paper: The Incidence of the Corporate Income Tax. March 1996.
- Congressional Budget Office Testimony: Trends in Federal Tax Revenues and Rates. Statement of Douglas Elmendorf, CBO Director, before the United States Senate Committee on Finance. Dec. 2, 2010.
- Dhaliwal, Dan, Cristi Gleason and Lillian Mills. “Last-Chance Earnings Management: Using the Tax Expense to Meet Analysts’ Forecasts.” *Contemporary Accounting Research* Vol. 21 No. 2. Summer 2004.
- Drucker, Jesse. “Google 2.4% Rate Shows How \$60 Billion Lost to Tax Loopholes.” Bloomberg. Oct. 21, 2010. <<http://www.bloomberg.com/news/2010-10-21/google-2-4-rate-shows-how-60-billion-u-s-revenue-lost-to-tax-loopholes.html>>

Exposure Draft ED/2009/2. Income Tax. Issued March 2009. International Accounting Standards Board.

Federal Accounting Standards Board. Current Technical Plan and Project Updates. <<http://www.fasb.org/cs/ContentServer?c=Page&pagename=FASB%2FPage%2FSectionPage&cid=1218220137074>> Accessed March 16, 2011.

Federal Accounting Standards Board Exposure Draft. Compensation – Retirement Benefits – Multiemployer Plans (Subtopic 715-80). Disclosure about an Employer’s Participation in a Multiemployer Plan. Issued Sept. 1, 2010.

Federal Accounting Standards Board Interpretation No. 48, Accounting for Uncertainty in Income Taxes, an interpretation of FASB Statement No. 109. June 2006. Financial Accounting Standards Board.

Statement of Financial Accounting Standards No. 109, Accounting for Income Taxes. Financial Accounting Standards Board. Issued February 1992.

Fiscal Year 2008 IRS Enforcement Results. Internal Revenue Service. <irs.gov/pub/irs-news/2008_enforcement.pdf>

Flora v. United States, 362 U.S. 145, 176 (1960).

General Accounting Office of the United States, Testimony before the Committee on Finance, U.S. Senate. “Internal Revenue Service: Challenges Remain in Combating Abusive Tax Shelters.” Statement of Michael Brostek, Director, Tax Issues. Oct. 21, 2003.

Gleason, Cristi and Lillian Mills. “Materiality and Contingent Tax Liability Reporting.” *The Accounting Review*. April 2002.

Gravelle, Jane and Thomas Hungerford. “Corporate Tax Reform: Issues for Congress.” Congressional Research Service. Oct. 31, 2007.

Hanlon, M., and J. Slemrod. 2006. What does tax aggressiveness signal? Evidence from stock price reactions to news about tax aggressiveness. University of Michigan working paper.

Helman, Christopher. “What the Top U.S. Companies Pay in Taxes.” *Forbes*. April 1, 2010. <http://www.forbes.com/2010/04/01/exxon-walmart-business-washington-corporate-taxes_2.html>

Higgins, Anthony. Statement. 26 Cong. Rec. 6874. 1894.

IASB Agenda Project. Convergence Issues – IFRSs and US GAAP: Income Taxes. Deloitte Global Services Limited.

Internal Revenue Manual. Internal Revenue Service.

Internal Revenue Service Advisory Council. Large Business and International Subgroup Report. Nov. 17, 2010. <irs.gov/taxpros/article/0,,id=231502,00.html>

Internal Revenue Service Announcement 2010-76.

Internal Revenue Service Data Release. “Corporation Income Tax Brackets and Rates, 1909-2002.”

Internal Revenue Service Notice 2002-50, “Tax Shelter.” Issued May 24, 2005.

Internal Revenue Service Notice 2005-83, “KPMG to Pay \$456 Million for Criminal Violations.” Issued Aug. 29, 2005.

International Accounting Standard 12, Income Taxes. As amended up to Dec. 31, 2009. Issued Oct. 1996. International Accounting Standards Committee Foundation.

International Accounting Standard 37, Provisions, Contingent Liabilities and Contingent Assets. As amended up to Dec. 31, 2009. Issued Sept. 1998. International Accounting Standards Committee Foundation.

Joint Committee on Taxation, *Background and Present Law Relating to Tax Shelters* (JCX-19-02), March 19, 2002.

Jones, Andrew. "FASB – The IRS's New Best Friend: How FIN 48 Affects the Taxpayer-IRS Relationship and Potential Taxpayer Challenges." *Georgia State University Law Review*, Vol. 25, Issue 3.

Kraw, George. Comment Letter to the Financial Accounting Standards Board. Sept. 15, 2010.

Kwiatek, Harlan. "FIN 48 and State and Local Income Taxes." Ed. Scott Salmon. *The Tax Adviser*. Sept. 1, 2007.

Lenter, David, Douglas Shackelford and Joel Slemrod. "Public Disclosure of Corporate Tax Return Information: Accounting, Economics, and Legal Perspectives." *National Tax Journal*. Dec. 1, 2003.

Levinsohn, Alan. "Shuttering Tax Shelters." *Strategic Finance*. Oct. 1, 2005.

Levitt, Arthur. "The 'Numbers Game.'" Remarks to the Securities and Exchange Commission at the New York University Center for Law and Business, New York, NY. Sept. 1998.

Lisowsky, Petro. "Seeking Shelter: Empirically Modeling Tax Shelters Using Financial Statement Information." Sept. 13, 2010.

McGill, Gary and Edmund Outlsay. "Did Enron Pay Taxes? Using Accounting Information to Decipher Tax Status." *Tax Notes*, Vol. 96, No. 8, Aug. 2002.

Miller, Steven. Directive for All Large Business and International Division Personnel. Sept. 2010.

Mills, Lillian. "Book-Tax Differences and Internal Revenue Service Adjustments." *Journal of Accounting Research*. 1998.

Mills, Lillian, Leslie Robinson and Richard Sansing. "FIN 48 and Tax Compliance." March 2009.

Montgomery, Lori. "Volcker-led economic panel pushes lawmakers to simplify U.S. tax code." *The Washington Post*. Aug. 27, 2010.

Pollock v. Farmers' Loan & Trust Company, 157 U.S. 429 (1895), aff'd on reh'g, 158 U.S. 601 (1895).

Potts-Dupre, David. Comment Letter to the Financial Accounting Standards Board. Sept. 20, 2010.

Public Law 86-272. 15 U.S.C. §381.

Report of the Commissioner of Corporations on the System of Taxing Manufacturing, Mercantile, Transportation, and Transmission Corporations in the States of Ohio, Indiana, Illinois, Michigan, and Wisconsin. Taxation of Corporations: Part III – Eastern Central States 3. 1911.

Robinson, John, Stephanie Sikes and Connie Weaver. "Performance Measurement of Corporate Tax Departments." Sept. 2009.

Schedule UTP Instructions, 2010. Internal Revenue Service, Department of the Treasury.

Shulman, Douglas. "Prepared Remarks of Commissioner of Internal Revenue Douglas H. Shulman before the Tax Executives Institute 60th Mid-Year Meeting." April 12, 2010.

"State Tax Nexus: Nexus, Multistate Tax and State Net Operating Loss Issues Practice Guides." AICPA. 2007.

Statement of Financial Accounting Concepts No. 1, Objectives of Financial Reporting by Business Enterprises. Financial Accounting Standards Board. Nov. 1978.

Statement of Financial Accounting Concepts No. 2, Qualitative Characteristics of Accounting Information. Financial Accounting Standards Board. May 1980.

Statement of Financial Accounting Concepts No. 6, Elements of Financial Statements. Financial Accounting Standards Board. Dec. 1985.

Statement of Financial Accounting Standards No. 5, Accounting for Contingencies. Financial Accounting Standards Board. March 1975.

Statement of Financial Accounting Standards No. 143, "Accounting for Asset Retirement Obligations." Financial Accounting Standard Board. June 2001.

Stecker, J. Howard. Letter of Comment No. 61, FIN 48. Sept. 12, 2005.

Stromsem, William. "Private Companies and FIN 48." *Journal of Accountancy*. Jan. 17, 2009.

Taft, William. Statement. 44 Cong. Rec. 3344. 1909.

Tax Policy Center: "The Numbers: What Are the Federal Government's Sources of Revenue?" April 22, 2009. Accessed Jan. 9, 2011.
<taxpolicycenter.org/briefing-book/background/numbers/revenue.cfm>

TEI Comments on IASB Income Tax Exposure Draft (IAS12). Vincent Alicandri, Tax Executive Institute. Aug. 7, 2009.

"The Constitution of the United States."

Treasury Department, "The Problem of Corporate Tax Shelters, Discussion, Analysis and Legislative Proposals." July 1999.

Tweedie, David. "The Future of Financial Reporting: Convergence or Not?" Address to the U.S. Chamber of Commerce. March 10, 2011.

United States of America v. Textron Inc. and Subsidiaries. United States Court of Appeals For the First Circuit No. 07-2631. 2009.

United States Code: Title 26, Subtitle F, Chapter 61, Subchapter B, §6103. "Confidentiality and disclosure of returns and return information."

U.S. Congress. House. Corporate Accountability Tax Gap Act of 2003. 108th Congress, 1st sess., 2003. H.R. 1556.

United States Department of the Treasury. "History of the U.S. Tax System."
<<https://ustreas.gov/education/fact-sheets/taxes/ustax.shtml>> Accessed Dec. 18 2010.

United States Senate Permanent Subcommittee on Investigations, Committee on Governmental Affairs. "U.S. Tax Shelter Industry: The Role of Accountants, Lawyers, and Financial Professionals." 2003.

United States v. Arthur Young & Co., 465 U.S. 805, Volume 465. 1984.

Wei, Ronald. "IRS Additional Guidance on Uncertain Tax Positions Still Concerns Financial Executives." *Financial Executive* Dec. 2010: 13-14.

Wessel, David. "Tax Redo Seeks 'Level Playing Field.'" *The Wall Street Journal*. Jan. 27, 2011.

Work Plan for IFRSs. Income Taxes. International Financial Reporting Standards Foundation. Accessed March 20, 2011.
<<http://www.ifrs.org/Current+Projects/IASB+Projects/Income+Taxes/Income+Taxes.htm>>

Yrjanson, Carla and Jeff Tuthill. "Leveraging the Knowledge of Complementary Tax Groups." *Financial Executive* Dec. 2010: 24-28.

Trans-identity in Djuna Barnes' *Nightwood*

Nicholas Becht

In her novel, *Nightwood*, Djuna Barnes presents the reader with Dr. Matthew O'Connor, a character whose identity is so richly contradictory and complex that he arguably has no fixed identity or true self. On the surface, he is a cross-dressing, homosexual gynecologist with no medical license who inserts himself in and mediates between the lives of the main characters in the novel. Although the narrator, the other characters, and he himself often associates him with homosexuality, Barnes never represents O'Connor engaging in a specifically homosexual sex act. Also problematic are his various acts of transvestism and/or cross-dressing, as well as his identification(s) with femininity and seeming desire to be or become a woman. He sometimes even expresses that he already is a woman, and other times refers to himself as "the girl that God forgot" (Barnes 73). Several sources help to make sense of—or, perhaps further complicate—Dr. O'Connor's identity.

Neil Miller's chapter, "Pioneers of Sexology," provides contextualizing information about the prevailing 'knowledge' and attitudes regarding non-normative gender identities and sexualities around the time that *Nightwood* was written and is set—1936, and the 1920s through the 1930s, respectively. Of particular interest is the notion of sexual inversion or of the soul being in the wrong body (Miller 14). Miller, through his use of Foucault, also makes relevant observations about the trend toward the categorization of identities by acts or actions in the late nineteenth and early twentieth centuries. Miller explains the medical categorization of the homosexual identity and explores its conflation with the feminine identity due to notions about sexual inversion in the nineteenth and early twentieth centuries:

Ulrichs assumed that because male homosexuals had a female soul in a male body, they therefore possessed the personality characteristics of women...according to him, homosexuality was not just an "inversion" in the choice of sexual object but an "inversion" of one's broader gender characteristics as well...his theory of the "third sex" gave these gender stereotypes a quasi-scientific basis, confounding sexual orientation with gender and homosexuals with hermaphrodites. (Miller 14)

Barnes reflects these attitudes in her characterization of Dr. O'Connor, who frequently remarks that he is, "just the girl that God forgot" (Barnes 73), "the bearded lady" (Barnes 100), "the Old Woman who lives in the closet" (Barnes 138), "the other woman that God forgot" (Barnes 143), and "a lady in need of no insults" (Barnes 151). It is interesting to note that while Dr. O'Connor makes these repeated identifications with women and femininity throughout the novel, he always uses masculine pronouns when referring to himself. One must wonder then, if he is using the popular and prevailing medical notions of the time about homosexuality, inversion, and "contrary sexual feeling" (Miller 13) in order to make sense of himself for himself and for others, or if he is merely expressing that which he feels to be his truest, most interior identity.

There are two major scenes in which Dr. O'Connor physically and visually expresses his conceptions and descriptions of himself both as a woman, as well as desiring to be or become a woman, through transvestism. In each scene, a different character—The Baron Felix first, and Nora Flood second—perceives, reacts, and responds to Dr. O'Connor's transgressions of gender. Following a deconstructionist approach, I will attempt multiple readings of each scene in order to further complicate Dr. O'Connor's identity, which arguably, given his existence in an authored fiction, is already unstable. I will also connect Felix and Nora's desire of and discomfort about Dr. O'Connor and his transvestism to Marjorie Garber's notion of the transvestite as "both terrifying and seductive," found in her writing, "Dress Codes, or the Theatricality of Difference." While the transvestite is generally held to be strictly haunting and unattractive, it is often the case that the transvestite is found to be in some ways very attractive, which contributes to an individual's sense of horror. Categories and identities are troubled when one discovers how much—or perhaps, how little—changes when as little as one signifying article is added or switched. Garber argues that the transvestite is, "a crisis of 'category' itself" (Garber 32). The following passage is Barnes' eroticized account of Felix observing Dr. O'Connor's transgressive actions in the hotel room of Felix's fiancée, Robin Vote:

Experiencing a double confusion, Felix now saw the doctor, partially hidden by the screen beside the bed, make movements common to the “dumbfounder,” or man of magic; the gestures of one who, in preparing the audience for a miracle must pretend that there is nothing to hide; the whole purpose of making the back and elbows move in a series of “honesties,” while in reality the most flagrant part of the hoax is being prepared. Felix saw that this was for the purpose of snatching a few drops from a perfume bottle picked up from the night table; of dusting his darkly bristled chin with a puff, and drawing a line of rouge across his lips, his upper lip compressed on his lower, in order to have it seem that their sudden embellishment was a visitation of nature...the doctor’s hand reached out and covered a loose hundred franc note lying on the table. With a tension in his stomach, such as one suffers when watching an acrobat leaving the virtuosity of his safety in a mad unraveling whirl into probable death, Felix watched the hand descend, take up the note, and disappear into the limbo of the doctor’s pocket. He knew that he would continue to like the doctor, though he was aware that it would be in spite of a long series of convulsions of the spirit, analogous to the displacement in the fluids of the oyster, that must cover its itch with a pearl; so he would have to cover the doctor. He knew at the same time that this stricture of acceptance (by which what we must love is made into what we can love) would eventually be a part of himself, though originally brought on by no will of his own. (Barnes 35-36)

Barnes’ choice of words is evocative and insinuating. Felix’s “double confusion” could be his reaction to Dr. O’Connor inhabiting the double, or dual role(s) of the transvestite as both male and female as well as “both terrifying and seductive” (Garber 32). Barnes emphasizes the performative aspects of Dr. O’Connor’s transvestism through her comparison of him to a magician and to an acrobat. The application of cosmetics themselves often precedes or signals a performance. It is interesting that Dr. O’Connor is described as preparing a “hoax” and attempting to have it seem that his suddenly rouged lips are a “visitation of nature,” when we take into consideration the fact that even when someone we consider to be or accept as a woman puts on perfume, powder, and rouge, she is also preparing a “hoax” and affecting a “visitation of nature;” Dr. O’Connor’s actions, which are

considered feminine, a “hoax,” and ‘unnatural’ for him to perform, are themselves deceptive, artificial, and ‘unnatural’ acts. These acts, as well as Dr. O’Connor’s theft of the hundred franc note, both “terrify” and seem to “seduce” Felix. Garber’s quotation and vivid summary of Dr. John Rainolds provides a possible explanation for what appears to be Felix’s uncomfortable attraction to and desire of Dr. O’Connor as the perfumed, powdered, and rouged thief:

For Rainolds, women’s clothes act as transferential objects, kindling a metonymic spark of desire: ‘because a women’s garment being put on a man doeth vehemently touch and moue him with the remembrance and imagination of a woman; the imagination of a thing desirable doth stir up the desire’ (Rainolds 96-97).’ (Garber 29)

In this case, Dr. O’Connor is not wearing women’s clothes, but a particular woman’s cosmetics and fragrance. That particular woman is Felix’s fiancée, Robin Vote, whose cosmetics and fragrance --one can imagine-- would kindle a rather large “metonymic spark of desire,” as well as a strong sense of discomfort and/ or “terror” for Felix. It is perhaps ironic to note that Robin is characterized by Barnes as being rather masculine, as well as bisexual and perhaps truly a lesbian. Given notions at the time of male and female sexual inversion, Dr. O’Connor’s appropriation of Robin’s identity through his appropriation of her cosmetics and fragrance strangely situates both his (sense of) identity as well as Felix’s desire(s) and (sense of) identity.

Felix’s feelings about and reactions to Dr. O’Connor’s transvestism and theft are described by Barnes in terms which are sensually and sexually suggestive, such as the image of Felix’s convulsing spirit, and the accompanying image of an oyster displacing its fluids in order to form a pearl. The word convulsing evokes the gyrations, thrusting, and orgasm which often accompany sexual intercourse and the formation of the pearl serves as a metaphor for ejaculation; the oyster itself is a gynic image. It is fitting then, that Felix would consider his desire and accompanying discomfort in terms of a convulsing spirit, or soul, rather than --or perhaps within-- a physical body, when confronted with homosexuality and/ or male sexual inversion. The “stricture of acceptance,” or rigid demands of a heteronormative society, is what in part inspires the concept of inversion, “By which what [Felix] must love is made into what [he] can love” (Barnes 36). Dr. O’Connor is a man who is made into a woman

both through his sexual inversion and through his transvestism, which renders Felix acceptably vulnerable to the “metonymic spark of desire” (Garber 29) which he justifies to himself as having been, “brought on by no will of his own” (Barnes 36). This locates Felix’s desire(s) within an unconscious mind, and perhaps within the soul of an unwitting male sexual invert.

While Felix witnesses part of the process of Dr. O’Connor’s transvestism which physically and visually represents –or, perhaps only hints at-- his gender transference and inversion, Nora encounters Dr. O’Connor at what seems to be the height of his transvestism:

In the narrow iron bed, with its heavy and dirty linen sheets, lay the doctor in a woman’s flannel nightgown. The doctor’s head, with its over-large black eyes, its full gun-metal cheeks and chin, was framed in the golden semi-circle of a wig with long pendent curls that touched his shoulders and falling back against the pillow turned up the shadowy interior of their cylinders. He was heavily rouged and his lashes painted. It flashed into Nora’s head: “God, children know something they can’t tell; they like Red Riding Hood and the wolf in bed!” But this thought, which was only the sensation of a thought, was of but a second’s duration as she opened the door; in the next, the doctor had snatched the wig from his head and sinking down in the bed drew the sheets up over his breast. (Barnes 79)

Dr. O’Connor uses his wig and other feminine signifiers to show his inversion and to turn up his “shadowy interior,” much in the same way that the curls of his wig show theirs’; in this way, he expresses the woman that he is and/ or identifies with. Nora’s immediate reaction comparing Dr. O’Connor to “Red Riding Hood and the wolf in bed” again reminds us of Garber’s notion of the transvestite as “both terrifying and seductive” (Garber 32), whereby he represents the dual role(s) of both the “seductive,” female Red Riding Hood as well as the “terrifying,” male wolf. Further troubling category and identity, Dr. O’Connor also represents, for Nora, the way(s) in which Red Riding Hood is fascinated, “terrified,” and “seduced” by the wolf’s transvestism as her grandmother, in her grandmother’s bed. Much in the same way as it does for Felix, Dr. O’Connor’s transvestism also troubles Nora’s (sense of) identity. Is Nora’s “sensation of a thought,” a brief, immediate observation, or is it a pleasurable, uncomfortable feeling like that which Felix

experiences in response to Dr. O’Connor’s transvestism? Nora is, however, unlike Felix in that she is a lesbian and/ or a female sexual invert. It would follow then that she might be “seduced” by Dr. O’Connor’s female interior and “terrified” by his still –despite his transvestism— overwhelmingly male exterior. The fact that she is Robin’s lover and is fascinated, “terrified,” and “seduced” by Dr. O’Connor’s transvestism again troubles his identity as well as her desire(s) and (sense of) identity.

Dr. O’Connor’s transvestism might also be considered cross-dressing, or drag. Esther Newton’s concepts about and explanations of the sartorial system which functions within drag – which she studied as an ethnographic anthropologist and wrote about in “Selection From Mother Camp”— prove useful in an analysis of the two, previously quoted scenes which feature Dr. O’Connor’s cross-dressing and/ or drag transgressions:

The principle opposition around which the gay world revolves is masculine-feminine... There are two different levels on which the oppositions can be played out. One is *within* the sartorial system itself, that is, wearing feminine clothing “underneath” and masculine clothing “outside.” ...It symbolizes that the visible, social, masculine clothing is a costume, which in turn symbolizes that the entire sex-role behavior is a role –an act...A second “internal” method is to mix sex-role referents *within* the visible sartorial system. This generally involves some “outside” item from the feminine sartorial system such as earrings, lipstick, high-heeled shoes, a necklace, etc., worn *with* masculine clothing...The feminine item stands out so glaringly by incongruity that it “undermines” the masculine system and proclaims that the inner identification is feminine...The second level poses an opposition between one sex-role sartorial system and the “self,” whose identity has to be indicated some other way. Thus when impersonators are performing, the oppositional play is between “appearance,” which is female, and “reality,” or “essence,” which is male...a drastic step is taking off the wig. (Newton 122)

Barnes’ description of Dr. O’Connor’s impressive collection consisting of “laces, ribands, stockings, ladies’ underclothing and an abdominal brace” (Barnes 78) suggests that he more than likely --and perhaps always— plays out the opposition of

masculine-feminine within the sartorial system and underneath his masculine clothing. He also mixes gendered signifiers within the visible sartorial system, such as when he puts on powder and rouge in front of Felix while wearing masculine clothing, or even when he receives Nora in almost full drag – except for his beard. In the first instance with Felix, however, feminine signifiers “undermine” the masculine system,” while in the second instance with Nora, a masculine signifier “undermines” the feminine system. The latter instance is similar to, but not quite the same as the second level oppositions which occur “between one sex-role sartorial system and the “self” (Newton 122). Dr. O’Connor does, however, perform this second level of opposition when he removes his wig in Nora’s presence, as well as when he presumably speaks to her in his male voice.

Dr. O’Connor’s repeated and varied drag transgressions further destabilize his identity both for his audience within Barnes’ novel – predominantly Felix and Nora—as well as for the audience, or readers, of Barnes’ novel. His speech, actions, and ‘hints’ become so absurd, contradictory, and muddled, that one might even wonder if Dr. O’Connor is perhaps a woman in drag as a man in drag as a woman, or perhaps has no knowable identity, after all. On the destabilizing qualities of drag with respect to identity, Newton states:

At the most complex, it is a double inversion that says “appearance is an illusion.” Drag says, “my ‘outside’ appearance is feminine, but my essence ‘inside’ [the body] is masculine.” At the same time it symbolizes the opposite inversion: “my appearance ‘outside’ [my body, my gender] is masculine but my essence ‘inside’ [myself] is feminine.”

(Newton 124)

It is Dr. O’Connor’s “double inversion,” or double role(s) as a homosexual or male sexual invert, who is also a transvestite, and in some ways performs drag transgressions, which causes Felix’s “double confusion” (Barnes 35) which is also experienced by Nora, as well as the readers of Barnes’ novel. Dr. O’Connor’s “double inversion” exists in the form(s) of his incongruous gender and sexual identities, and the resulting “double confusion” is merely a perception of and reaction to those incongruities.

Newton examines two related, though separate manners, in which such incongruities are addressed and suggests that “The drag queen simply expresses the incongruity while the camp actually uses it to

achieve a higher synthesis” (Newton 125). She then defines camp as “the tension between that person or thing and the context or association” (Newton 126). Both in Barnes’ presentation of Dr. O’Connor and in Dr. O’Connor’s presentation of self –or selves—he appears to achieve some form of this “higher synthesis.” Dr. O’Connor and his incongruous signifiers of gender and sexuality both in the context of Robin’s hotel room and his association there with Felix, as well as in his bedroom and his association there with Nora, exhibit a tension which can be considered camp. The former example can even be considered “unintentional camp” (Newton 125), in that Dr. O’Connor is either not aware that Felix is witnessing his transgressions or truly believes that he is deceptive enough to commit his transgressions without drawing any attention to himself. The latter example exhibits the transformative quality of camp. Newton explains, “Camp humor is a system of laughing at one’s incongruous position instead of crying. That is, the humor does not cover up, it transforms.” (Newton 127). When Dr. O’Connor observes Nora’s shock and discomfort at having found him dressed and made up as a woman, he jokes, “You see that you can ask me anything,” thus laying aside both their embarrassments” (Barnes 80). He both acknowledges and laughs, thus inviting Nora to laugh, at his gender incongruity, and in doing so transforms the context of and removes the tension from their situation.

Drawing from her ethnographic experience and observations, Newton also adds: “I saw the reverse transformation –from laughter to pathos—often enough and it is axiomatic among the impersonators that when the camp cannot laugh, he dissolves into a maudlin bundle of self-pity” (Newton 127).

In Dr. O’Connor’s final scene, he comments in a drunken tirade on the nature of camp humor in a way that is maudlin and –perhaps unintentionally– campy: “Only the scorned and ridiculous make good stories,’ he added angrily, seeing the habitués smiling, ‘So you can imagine when you’ll get told! Life is only long enough for one trade; try that one!” (Barnes 159). In a way, camp humor is a story about the scorned and ridiculous told by the camp figure, Dr. O’Connor, who –through his sexual inversion, transvestism, drag transgressions, and expressions of camp-- is and renders himself both scorned and ridiculous: A grotesquely fascinating spectre of gender and sexual incongruity.

Works Cited

Barnes, Djuna, and T. S. Eliot. *Nightwood*. New York: New Directions Pub., 2006. Print.

Garber, Marjorie B. *Vested Interests: Cross-dressing & Cultural Anxiety*. New York: Routledge, 1992. Print.

Miller, Neil. *Out of the Past: Gay and Lesbian History from 1869 to the Present*. New York: Alyson Books, 2006. Print.

Newton, Esther. "selection from Mother Camp." Stryker, Susan, and Stephen Whittle. *The Transgender Studies Reader*. New York: Routledge, 2006. Print

Cat on a Hot Streetcar Named Desire: Interrogations of Femaleness—Or, the Mad Heroine Coded as Homosexual—in Two Plays by Tennessee Williams

Nicholas Becht

Tennessee Williams utilizes heterosexual female protagonists in two of his major plays, “A Streetcar Named Desire” and “Cat on a Hot Tin Roof,” in order to covertly represent and explore themes of male homosexuality and desire. They are Blanche DuBois and Maggie Pollitt, respectively. He does this because he is writing for a largely conservative and heterosexual audience of which anyone who might be homosexual would likely be in the closet. This is evocative of Eve Kosofsky Sedgwick’s notion of the “Open Secret” (Sedgwick 52).

Through various means, the identities of these supposedly heterosexual women are queered. Through their involvement and presumably sexual partnerships with men of dubious heterosexuality, Blanche DuBois and Maggie Pollitt become aligned with homosexuals and associated with homosexuality. Williams employs elements such as Camp and subverted forms of the Patriarchal Gaze in order to further interrogate and trouble our sense of Blanche’s and Maggie’s femaleness.

While describing her late husband, Allan Grey, and his suicide, Blanche makes allusions to his homosexuality:

There was something different about the boy, a nervousness, a softness and tenderness which wasn’t like a man’s, although he wasn’t the least bit effeminate looking --still—that thing was there.... He came to me for help. I didn’t know that. I didn’t find out anything till after our marriage when we’d run away and come back and all I knew was I’d failed him in some mysterious way and wasn’t able to give the help he needed but couldn’t speak of! He was in the quicksands and clutching at me – but I wasn’t holding him out, I was slipping in with him! I didn’t know that. I didn’t know anything except I loved him unendurably but without being able to help him or help myself. Then I found out. In the worst of all possible ways. By coming suddenly into a room which I thought was empty –which wasn’t empty, but had two people in it...the boy I had married and an older man who had been his friend for years.... (Williams 95)

The words that Blanche uses to describe her husband –“nervous,” “soft,” “tender”—are all

words which are used to describe her. His coming to her for “help” parallels her own helplessness and looking to others for “help.” Their shared qualities and helplessness along with the imagery of her slipping into the quicksand with him indicates Blanche’s descent into and association with homosexuality. In some way Allan remains attracted to Blanche, and she remains attracted to him. She is complicit with his homosexuality. This is exhibited in her driving out to the casino with him and his old friend. Her placement in this love triangle makes her doubly associated with homosexuality.

In “Cat on a Hot Tin Roof,” Maggie remains married to Brick Pollitt, a man who will no longer make love to her and who is tortured by the death of his friend Skipper with whom Maggie had an attempted but unsuccessful sexual liaison. Allusions to Brick and Skipper’s homosexuality are made throughout “Cat on a Hot Tin Roof.” Big Daddy and Maggie are perhaps most confrontational with respect to Brick. Maggie reflects candidly to Brick:

I said, “SKIPPER! STOP LOVIN’ MY HUSBAND OR TELL HIM HE’S GOT TO LET YOU ADMIT IT TO HIM!” –one way or another! HE SLAPPED ME HARD ON THE MOUTH!—then turned and ran without stopping once, I am sure, all the way back to his room at the Blackstone....—when I came back to his room that night, with a little scratch like a shy little mouse at his door, he made that pitiful, ineffectual attempt to prove that what I had said wasn’t true...—in this way, I had destroyed him, by telling him the truth that he and his world which he was born and raised in, yours and his world, had told him could not be told? (Williams 42- 43)

Maggie and Skipper’s “attempt” confirms both Skipper and Maggie’s –and perhaps even indirectly, Brick’s-- homosexuality. Skipper’s inability to perform sexually with a woman implies a sexual attraction to men, whereas Maggie’s attempt at intercourse with a man who is attracted not to women but men suggests that her desires too, are perhaps queer or misplaced. Both Maggie and Skipper evidently make this attempt in order to feel closer to Brick, which establishes another strange

love triangle. Maggie, like Blanche is also doubly associated with homosexuality.

By association and allusion Blanche and Maggie both become something like the homosexuals with whom they are involved. Why are these particular men attracted to these particular women, and why are these particular women attracted to these particular men? This relational disconnect suggests something about the nature of Blanche's and Maggie's appeal, attractions, and desires; in this way, we may regard them as homosexuals. Their attractions to and desires – though, likely unconscious—to be with homosexuals are homosexual. There exists an element of homosexuality in their minds and psyches. It is almost as if they are reconciled, yet still subverted or perverse versions of Kraft-Ebbing's invert. They are men who are attracted to and desire men, living in the bodies of women.

Blanche and Maggie are also characterized in manners similar to their respective homosexual partners. They are all described as being dishonest, "weak," "sick," and in some way inadequate. Allan is "degenerate" and Skipper and Brick are both alcoholics. Blanche drinks. Skipper is also "crippled" and needs his crutch. Maggie is driven wild by her unfulfilled desires. While the men lie to hide their sexuality, the women also lie about and conceal other aspects of their lives, including but not limited to their sexuality.

Blanche is unable to keep her family's plantation, Belle Reve, and her job as a High School English teacher on account of her "weaknesses," which appear to be men –especially young boys (particularly a seventeen-year-old student)—and clothing and jewelry. Stanley suspects that she is trying to "swindle" him out of what he deserves by the Napoleonic Code, but what he might also subconsciously suspect is that she is trying to "swindle" him into believing that she is a lady.

Maggie comes from a poor family background and as a result, feels inferior and defensive. She is also deemed and feels inadequate as a wife and hopeful mother since her husband will neither make love to nor impregnate her.

Another way in which Blanche and Maggie are situated as being more like homosexual males than heterosexual females is through the interrogation of their femaleness and exaggeration of their femininity.

Blanche's aging, fading beauty, failed marriage to Allan Grey, and failure to become engaged to Harold Mitchell suggest her failure and

illegitimacy as a woman. Marriage is the ultimate reification of heteronormative values. Blanche's failed marriage to a homosexual renders her illegitimate as a woman in the eyes of society both in that it actually failed and that it was, or became, a homosexual relationship. Blanche's status as a single woman and seeming inability to become engaged and remarried is a threat to heterosexuality and heteronormativity. Heteronormativity dictates that heterosexuality can only be achieved and known through the production of the traditional gender binary by two persons of the opposite sex -- a man and a woman—who must also fulfill specific, gendered roles.

Blanche relies on stereotypical signifiers of femininity in order to (re)assert herself as a "real" or "true" woman, admitting, "I know I fib a good deal. After all, a woman's charm is fifty percent illusion" (Williams 41). This statement suggests what Judith Butler calls the "Performativity" of gender. Blanche is consummate in her illusion and performance of femininity. She bathes and primp often. She wears makeup, perfume, and stylish clothing and jewelry. She sings songs, swishes and dances around. She smiles, laughs, jokes, and flirts. She recites poetry and speaks in French. She talks about her nerves and overemphasizes her frailness, vulnerability, and femininity. She cannot even stand to be in the light of day or a naked light bulb, which speaks to the sort of deception she is attempting achieve. Mitch confronts her by saying, "I don't think I ever seen you in the light," and "I've never had a real good look at you" (Williams 116). Though superficially he is concerned with her true age and appearance, there is almost a sense that he is challenging her femaleness and wondering whether or not she is really a woman.

Blanche defends herself: "I don't want realism. I want magic! Yes, yes, magic! I try to give that to people. I misrepresent things to them. I don't tell the truth, I tell what *ought* to be truth. And if that is sinful, then let me be damned for it!" (Williams 117). Her statement about what ought to be the truth suggests that it ought to be permissible for two men to be in love and in a relationship with each other. This idea is strengthened by Blanche's mention of sin and damnation with respect to her (homosexual) behavior.

Gordon Rogoff posits that "Blanche is not a woman at all, she is a man in drag" (83). Quentin Crisp explains, in an interview, the ways in which effeminate homosexuals attempt to win and keep the love and affection of "real" men through the reproduction of heteronormativity, gender binaries,

and stereotypes: "They simply become more feverish, they pile the makeup on thicker, they sway more from side to side, they Camp more outrageously, and of course, the more they do this, the more the desired object recedes."

Blanche is of course feverish and only becomes more so throughout the rise of action in "A Streetcar Named Desire." She is anxious in her performance of femininity and in her seductions and attempts to ensnare a second husband, but as Crisp predicts, the desired object and objects recede. Her self-consciousness about the act makes others aware.

Crisp's use of the term 'Camp' reminds one of Susan Sontag's "Notes" on the subject. Blanche is a "Woman," "in quotation marks" (Sontag 109). She possesses "a relish for the exaggeration of sexual characteristics and personality mannerisms...corny flamboyant femaleness" (Sontag 109). Her "character is understood [by the other characters and the audience] as a state of continual incandescence – a person being one, very intense thing. This attitude toward character is a key element of the theatricalization of experience embodied in the Camp sensibility" (Sontag 114). How fitting that the actress playing Blanche also plays Blanche playing a role as herself.

Sontag also explains "camping:" "the vulgar use of the word Camp as a verb, 'to camp,' something that people do...a mode of seduction – one which employs flamboyant mannerisms susceptible to a double interpretation; gestures full of duplicity," and continues, stating, "One must always distinguish between naïve and deliberate Camp. Pure camp is always naïve. Camp which knows itself to be Camp ('camping') is usually less satisfying. The pure examples of Camp are unintentional; they are dead serious" (110).

While there may be some elements of pure, naïve, unintentional Camp to Blanche's personality, the majority of her behavior –especially with respect to men—is 'camping.' Blanche is at times, certainly duplicitous. Also, Stanley and later, Mitch, grow tired and dissatisfied with, and begin to see through Blanche's 'camping.' One might suppose, however, as the dialogue indicates, that Blanche began life as a very sweet and innocent, loving girl.

Blanche's brother-in-law, Stanley Kowalski, is most acutely aware of her deceptions, stating plainly:

I've been on to you from the start! Not once did you pull any wool over this boy's eyes!
You come in here and sprinkle the place with

powder and spray perfume and cover the light bulb with a paper lantern, and lo and behold the place has turned into Egypt and you are the Queen of the Nile! (Williams 127-28)

In essence, Stanley is telling Blanche exactly what kind of low woman he believes her to be. Taking the idea a bit further, he might be implying that she is not a woman at all, or at the very least, not a lady.

Ironically, Stanley still becomes sexually associated with Blanche, through raping her. To some extent, even her mimicry and parody of femininity is attractive –perhaps repulsively so-- to a presumably thoroughly heterosexual male.

Although Maggie is described as beautiful, sexy, and desirable to almost every other man besides Brick, her seeming inability to sexually please her husband and produce children are often referenced and questioned. Big Mamma asks her, "D'you make Brick happy in bed?", and tells her, "Something's not right! You're childless and my son drinks!" (Williams 32-33). Mae later remarks, "Do you know why she's childless? She's childless because that big beautiful athlete husband of hers won't go to bed with her!" (Williams 121). These "flaws" serve to destabilize her identity as a woman. The imagery as Brick as "big" and "beautiful" also serves to make him seem like less of a man and perhaps more like a maternal, matronly, or pregnant woman.

The fact that Brick does not desire and is not sexually satisfied by Maggie says just as much about Maggie's desirability as it does about Brick's desires. The fact that she has no children and is not pregnant suggests to the other characters who criticize and scrutinize her that she is somehow less of, lacking as --or perhaps, not-- a woman.

Blanche and Maggie are also more traditionally male, or masculine, in their desires. They frequently objectify men through compliments and their use of the gaze. To deliver the gaze is traditionally a male behavior, whereas to be the object of the gaze is a traditionally female position. This is especially evident and apparent in the film adaptations of both plays.

The stage directions in "A Streetcar Named Desire" indicate that Blanche "regards a [young man] with interest" (Williams 82). She looks at, touches, and makes him uncomfortable, telling him that he looks "like a young prince out of the Arabian Nights" and saying, "I want to kiss you, just once, softly and sweetly on your mouth!" (Williams 84). In a show of dominance, she kisses him without his consent and then asks him to leave.

The 1951 film version includes long shots of the young man. The camera overemphasizes the gaze. Also, the young man is shown to be pretty, almost like a girl, with big, feminine eyes and eyelashes.

Blanche's sister, Stella explains to her husband, Stanley, that Blanche objectified her late husband, Allan Grey: "He was extremely good-looking. I think Blanche didn't just love him but worshipped the ground he walked on! Adored him and thought him almost too fine to be human!" (Williams 102). This championing of beauty is typically something that males do with respect to females and not the other way around. Blanche has rendered Allan an object and therefore not human; she has dehumanized him. That this tribute was paid not only to a man, but to a male homosexual further complicates the notions of heteronormativity and stereotypical, binaried gender roles. In making Allan less of a man and more of a woman, Blanche makes herself more of a man and less of a woman. This and "Cat on a Hot Tin Roof," are both plays about queer sexualities: feminine masculinity and masculine femininity.

Throughout "A Streetcar Named Desire," Blanche tends to objectify and dehumanize Stanley as she characterizes him not as a man, but as an animal. She calls him an "ape" and "primitive" and considers him to be only sexual. She later refers to him and Mitch as "swine" undeserving of her fine company (Williams 126).

In "Cat on a Hot Tin Roof," Brick is frequently the object of Maggie's gaze. She often calls him "Baby" and other pet names and remarks that he is still beautiful and has kept his figure despite having taken to drink. Maggie also speaks about keeping her figure and mentions that other men still look at her. That she and Brick are both objects of attention places them together in an interesting way. She is constantly forcing herself on him and trying to persuade him to make love to her. In further reversal of the heteronormative gender roles, he is constantly refusing her and fighting her off. In the 1958 film adaptation, we are shown great shots of Liz Taylor's eyes as she slowly, deliberately, and unashamedly sizes up Paul Newman; she is not a timid woman and appears to feel entitled to get as good a look for as long as she likes.

Much like Blanche has done with Allan Grey, Maggie objectifies Brick through worship. She speaks of his desirability and unavailability:

Skipper and I made love, if love you could call it, because it made both of us feel a little bit closer to you. You see, you son of a bitch, you

asked too much of people, of me, of him, of all the unlucky poor damned sons of bitches that happen to love you, and there was a whole pack of them besides me and Skipper, you asked too goddamn much of people that loved you, you –superior creature! –you godlike being! –And so we made love to each other to dream it was you, both of us! (Williams 40)

Both women have strong desires, and Maggie especially tends to reject traditional notions of female morality and purity in favor of traditional "manly" pleasure-seeking. There appears to be an unsuccessful struggle either to become wholly female or to subvert what it means to be feminine.

Ultimately, it is the desire for men, or the way in which it is pursued, that results in dissatisfaction, anxiety, and loneliness for these women who are coded as homosexuals. This speaks to the loneliness of homosexuality, especially during the time in which Williams was writing.

Concerning "A Streetcar Named Desire," but also relevant to "Cat on a Hot Tin Roof," William Sharp observes, "Sex is comfort in the play, the one solace to all one's fears, and to all one's guilts as well" (162). Maggie can't get sex and practically goes insane, and Blanche is raped, and actually does go insane. There is also Stanley's –and Stella's, too, by association—guilt surrounding Stanley's rape of Blanche. Sex and lack of sex are both imbued with fear and guilt. Sex may make one feel close to another but is often not enough. Blanche claims to prefer matters of the mind, but clearly desires attention and places value on the physical. Brick values his "pure" friendship with Skipper over his sexual relationship with Maggie.

On Homosexual presence and themes in the plays of Williams, Gordon Rogoff writes:

The crisis of veiled homosexuality is, then, one of those decoys invented by a critical fraternity, which is itself in crisis. The plays, singly and together, are nothing if not revealing. What is revealed has something, perhaps a good deal to *do* with homosexuality, but it has nothing to *show* of the deed itself, which is hardly surprising. Drama is not yet documentary, despite all the ill-conceived efforts of newspaper naturalists with their well-made plays. On the contrary, drama selects from experience and transforms it into something else, something with a substance, order, and life all its own...One clinical interpretation could, with some justice, lean again toward homosexuality. To view life from a homosexual bedroom, cluttered as it is

with the memories of scattered partners, angry justifications, and broken dreams, is to view life as a prison. One wishes to be plural, but one is continually, and often agonizingly single. That is generally a simple, not easily changeable fact, of homosexual life. The mind thinks it is calling for one thing while its sexual surrender is moving it toward another. (85-86)

Maggie laments, "Living with someone you love can be lonelier –than living entirely alone! If the one that y' love doesn't love you" (Williams, 16), and Blanche exclaims, "I *can't* be *alone!*" (Williams 22).

The homosexuality about which Williams writes is a doomed and lonely homosexuality. It is a homosexuality doomed by its attempt to reproduce heterosexuality and the heteronormative gender binaries which do not even work in heterosexual relationships as is evidenced by the troubled heterosexual relationships in the plays: Stella and Stanley, Eunice and Steve, Big Daddy and Big Mamma, and Gooper and Mae. It is a homosexuality in which there still exists a dichotomy of power and desire that cannot be reconciled and leaves its participants dissatisfied and lonely if they are lucky and makes them go insane if they are not.

Works Cited

- Cat on a Hot Tin Roof*. Dir. Richard Brooks. Perf. Elizabeth Taylor, Paul Newman, and Burl Ives. DVD. MGM. 1958.
- Cleto, Fabio. "Introduction: Queering the Camp."
- Jagose, Annamarie. *Queer Theory: An Introduction*. New York: New York University Press, 1996.
- Parker, Robert Dale. *How to Interpret Literature: Critical Theory for Literary and Cultural Studies*. New York: Oxford University Press, 2008.
- Quentin Crisp*. Dir. Denis Mitchell. Perf. Quentin Crisp. 1970.
- Rogoff, Gordon. "The Restless Intelligence of Tennessee Williams." *Tulane Drama Review* 10.4 (1966): 78-92. Web. 20 April 2010. <<http://www.jstor.org/stable/1125211>>.
- Sedgwick, Eve Kosofsky. "Epistemology of the Closet."
- Sharp, William. "An Unfashionable View of Tennessee Williams." *Tulane Drama Review* 6.3 (1962): 160-171. Web. 20 April 2010. <<http://www.jstor.org/pss/1124941>>.
- Sontag, Susan. "Notes on 'Camp'."
- A Streetcar Named Desire*. Dir. Elia Kazan. Perf. Vivien Leigh and Marlon Brando. DVD. Warner Bros. 1951.
- Williams, Tennessee. *Cat on a Hot Tin Roof*. New York: Signet, 1983.
- . *A Streetcar Named Desire*. New York: Signet, 1975.

Defined by the Home: Housing and Class Connections in George Orwell's *Keep the Aspidistra Flying*

Sarah Graham

In George Orwell's *Keep the Aspidistra Flying*, 1930s class stratification in London, England takes center stage. Due to housing shortages after WWI, slum life became more prominent in the eyes of every Briton. Inadequate government funding lead to overcrowding and the regression of tenement housing conditions, which only resulted in these lower-class houses falling into greater ruin. Thus, differences between the middle and the growing lower-class, especially in regards to housing, became more and more pronounced, until housing alone could be a determinant of class. Also due to the housing shortages, the working classes were unable to escape their living situations, and thus became influenced by the ruin that surrounded them. Using *Keep the Aspidistra Flying* as the primary text, I will attempt to show how housing and class are dependent on each other in 1930s London, especially in regards to the exclusive nature of class, the size of personal space, ownership of personal space, levels of health, and freedom of sexuality.

I focus entirely on the two main couples of *Keep the Aspidistra Flying*. First are the upper-class characters of Ravelston and Hermione who live in luxurious flats. The other couple is the lower-class Gordon and Rosemary, who live in rented rooms in tenement houses. Between these couples, a definite separation can be witnessed, due entirely to their housing differences. Gordon is a poor bookshop clerk who feels ashamed of his squalid living situation. Even though he and Ravelston are good friends, Gordon proclaims that he "never, if he could help it, set foot in Ravelston's flat. There was something in the atmosphere of the flat that upset him and made him feel mean, dirty and *out of place*" (emphasis added 79-80). Gordon names a specific reason why he cannot enter Ravelston's home: it's because it is "overwhelmingly, though unconsciously, upper-class" (80). Ravelston is a magazine editor from a wealthy family, and his home reflects his fortune. Gordon feels he doesn't belong in that rich setting, which creates a solid divide between Ravelston and Gordon, which relates to their classes as a whole. Gordon "felt he had no business [in Ravelston's flat]- that this wasn't the sort of place where he belonged. There

was a sense of guilt [...] when he was ruined and hadn't a penny in the world" (187). Because of the evident class differences, Gordon is not at ease or accepted in an upper class home.

Conversely, when Ravelston visits Gordon's home and encounters what working-class housing is like first-hand, he thinks it is "dreadful to think of anyone with brains and refinement living in a place like this," thus insinuating that the lower-classes do not have either brains nor refinement (210). Ravelston briefly visits Gordon in the slums, but "the smelly place oppressed him," and he quickly leaves (211). As an upper class citizen, the homes of the lower-class are not where Ravelston belongs. Sanitary inspectors in London in the 1930s recorded that "large numbers of [London] citizens [were] huddled together in one- or two-roomed dwellings in a most unhealthy, immoral, disgraceful, and degrading manner [...] *Swine live better*." (emphasis added, Quigley 92, 94). If the poor are forced to live in such conditions because of their economic situation, they are forced to become part of that class, and mentally feel subordinate to upper-class citizens. One school of thought regarding slum tenants says that it is "senseless to say or to expect the slum hovel to produce a class compounded entirely of persons of high moral quality or heroic caliber," therefore living in slum and tenement housing produces lower-class citizens (Quigley 133). It creates a vicious cycle.

Orwell's depiction of ownership of personal space suggests that this is part of what defines stratified class housing. To begin with, we are presented with Gordon's first housing arrangements: a lower-middle-class bedsit home on the edges of the slums. He rents his small, dirty, drafty room from his overbearing-mother-like landlady Mrs. Wisbeach. Gordon pays for Mrs. Wisbeach to tidy his room, cook his meals, and probably do his laundry, like some sort of mother for grown men. To stay in her lodgings, Gordon must follow her rules, in which "tea-making [is] the major household offense, next to bringing a woman in" (29). Though Gordon regularly breaks the tea rule, he must do so in absolute secrecy, for Mrs. Wisbeach has been known for "sneaking upstairs and catching you in the act" (29). Even letters,

personal objects delivered from the world outside the lodging house, are “pawed about” by Mrs. Wisbeach before being delivered (33). Gordon obviously has no freedom in his house, for he does not own the building nor his room. The amenities are rented to him, and he must toe the line for as long as he stays there. Gordon doesn't even own the furniture in his homes; they are rented from the landladies.In fact, neither Rosemary nor Gordon “had ever owned furniture before; they had been living in furnished rooms ever since childhood” (245). Gordon explains that there is no “mingy lower-middle class decency [in the slums],” for the buildings and the people are so run down that it is impossible to pretend you’re not the lowest class (207). Gordon's lack of ownership of anything around him; his room, furniture, or the building, traps him within the confines of lower-class.

Contrasted to Gordon’s housing is Ravelston’s living arrangement. While Gordon lives in one-room “homes,” Ravelston lives alone in a “four-roomed flat, which he thought was a poky little place” (80). He does not have a landlady and is free to do whatever he pleases. Attention must also be drawn to the furniture in Ravelston’s house, all of which he owns, as opposed to renting it like Gordon. Gordon is given a chance to try out this furniture, as he “awoke [after being released from prison] in a wide caressing bed, softer and warmer than any bed he had ever slept in” (186). Gordon has never owned anything as nice as this bed, and probably never will, because of his income and class. Ravelston's freedom and luxury through ownership, due to his thick pocket book, is totally opposite to Gordon's situation.

If we again look back at Orwell’s housing portrayals, the dirt and disease in tenement housing becomes increasingly apparent. Orwell describes in detail Gordon's slum housing by saying:

[Gordon] never made his bed properly, but just turned back the sheets, and never washed his few crocks till all of them had been used twice over. There was a film of dust on everything. In the fender there was always a greasy frying-pan and a couple of plates coated with the remnants of fried eggs. One night the bugs came out of the cracks and marched across his ceiling two by two. (208)

Hermione, as an upper-class citizen, is also quick to attach filth to the stigma of being poor, proclaiming (about the lower classes), “I hate them. They *smell*” (93). This is a justified opinion as seen by the above description, and by the explanation that when Gordon moves to the heart of the slums, he “only

washed the parts that showed” (208). Even a Medical Officer of London comments in a 1930s medical report that “[i]n view of the slur often cast upon residents in the East-end [the worst slums] concerning their alleged dirty and verminous conditions [...] there can be no doubt that [...] some dirt and vermin is inevitable” (Quigley 135).

Beyond space, ownership, and cleanliness, each character’s class also controls their sexual freedom. Gordon and Rosemary, because of their lower-class housing, cannot have sex, for other people control their personal space. When Rosemary visits Gordon she is “never allowed indoors, not even into the hall,” because Mrs. Wisbeach is adamant about women not entering her abode. Mrs. Wisbeach has a sexual control over Gordon, both literally and figuratively, because of her ownership of the house. Similar to Gordon's situation, Rosemary lives “in a women's hostel” that is run by “she-dragons,” where males are not allowed in (94, 121). Therefore, the two of them must leave their houses, and reject the city and all forms of housing altogether, in order to have intimacy. They go to the countryside in order to have sex under a bush, like wild animals, thus degrading them and forcing them to feel even more like the lower class.

Rosemary especially is concerned with what could result from unprotected sex, for she will lose her job if she becomes pregnant. If she loses even the small income she has now, she would be unable to survive. She acknowledges the alternative, where she can “[have] it done for only five pounds” (226). Unfortunately, the abortion Rosemary is referring to is probably an unsafe one, performed by amateur doctors, or by people who are not doctors at all. This demonstrates how class controls housing and thus also controls sex. Even in abortion practices, class is apparent. Pamela Graves explains:

Like birth control, abortion was a practice to which women of all classes resorted [even though it was illegal], yet one where class discrimination clearly prevailed. Women who could afford to pay for the service were able to secure abortions secretly from sympathetic doctors [...] working-class women, on the other hand, had to rely on neighbours who ‘helped out’ [...] or [on] self-induced abortion. Both methods involved a higher incidence of death and injury than among better off women. (196)

This was at first an unrecognized pattern, but declining population prompted the government to look over maternal mortality studies, which

inevitably called attention to class-divided abortion-related deaths. Gordon and Rosemary cannot afford condoms or safe abortions, thus reflecting the dangers their class forces upon them when it comes to sexual freedom.

The other couple in the novel, Ravelston and Hermione, use the privacy their class provides them in order to enjoy sexual freedom. Ravelston has no landlady, and therefore, no one to screen who enters his home. Hermione has her own key and can come and go as she pleases. While Ravelston returns home at night to his waiting female, Gordon must return to “his foul lonely room [...] his womanless bed” (95). Ravelston and Hermione have been “lovers two years,” yet have never gotten married or lived in the same home (93). It is obvious from their casual attitude that they, especially Hermione who is described as ‘rich,’ can afford free sexuality, including birth control and safe abortions if necessary.

At the end of the novel, Gordon and Rosemary channel Orwell's personal views when they finally escape their lower-class status by upgrading to a private apartment (after Gordon gets a higher paying job). They get married and move to a flat that is out of the slums. They do not share the flat with anyone else; “Oh Gordon,” says Rosemary, “what fun it all is! To have a place that's really our own and no landladies interfering” (246). They “fell into absurd raptures over each separate stick of furniture,” for they even owned the beds and chairs they sat on, instead of renting them like they had their whole lives (245). This demonstrates their growth and ownership of personal space, and makes them consumers, which they had not experienced before. They even use part of their income to right away buy an aspidistra for themselves, the ultimate symbol of the middle-class. Their change of housing and acquisition of belongings boosts them into a higher class, which then allows them the privacy and resources to have sex, and deal with the results of sex (pregnancy).

As I have demonstrated, housing in London in the 1930s made huge gaps between the rich and

poor more apparent, and resulted in class cultures that were defined by housing. Lack of ownership of personal space, lack of privacy, lack of sanitation, and lack of sexual freedom all defined the lower-classes and was directly connected to their housing arrangements. Much like the deterioration Gordon succumbs to as the novel progresses, the London City Council proclaimed in their 1937 London housing report that the “psychological effect of living in mean surroundings, in houses which are dark, damp and dilapidated, and where privacy and cleanliness are obtained with difficulty, if at all, cannot be neglected in any attempt to assess the effect of faulty environment on the mental, moral, and physical fibre of the occupants” (London Housing 13). Tenement housing made the lower classes think and feel like lower class, and they were forced to accept this and stay within its confining nature.

Works Cited

- Orwell, George. *Keep the Aspidistra Flying*. New York: Harcourt, Brace, 1956. Print.
- Gater, G.H. *London Housing*. London City Council, 1937. Print.
- Graves, Pamela M. *Labour Women; Women in British Working Class Politics 1918-1939*. Cambridge UP, 1994. Print.
- "The Health of London: Report For 1931." *The British Medical Journal* 2.3744 (1932): 677-78. Print.
- Mowat, Charles Loch. *Britain between the Wars*. London: Methuen, 1972. Print
- Quigley, Hugh, and Ismay Goldie. *Housing and Slum Clearance in London*. London: Methuen & Co., 1934. Print.

A World Apart: Silko's Support of a Gynocentric Existence in *Gardens in the Dunes*

Sarah Graham

Leslie Marmon Silko is a Native American feminist and author of the 1998 novel *Gardens in the Dunes*, which focuses on the contradictions between repressed females in the patriarchal Western culture and self-determined Native American women. Her skillful storytelling expresses the power that can be drawn from being part of “collective sisterhood” that resists existing within the confines defined by patriarchy. Silko’s representation of repressive upbringings under patriarchy, positive portrayal of self-determined sexuality, and assertion of the existence of a “collective sisterhood” outside the influence of Western culture lays the foundation for Silko’s stance of challenging accepted gender roles in a mainly patriarchal world.

In *Gardens*, Silko introduces us to Hattie Palmer, an intelligent upper class white woman stuck in a patriarchal world in the early twentieth century. Hattie is the protagonist whom Silko deliberately traps in a patriarchy-defined world, so that we may watch her collapse under its influence and then rise above its restrictions. Silko first focuses on the approach of the education of women during this time period, an ironic commentary when considering the accepted notion that learning is a liberating experience. Hattie’s father initially supports her avid learning and she completes her early education “under the direction of her father” (Silko 95). Hattie then enrolls in university schooling, at a university run by men. Upon presenting her thesis, “The Female Principle in the Early Church,” she is immediately turned down by the male thesis review committee because this female perspective is deemed “a peripheral detail” (101). A male friend at college, Mr. Hyslop, compares her thesis “to the ‘lofty and spiritual ambitions’ of Margaret Fuller,” passing off her hard work as a joke and refusing to take her seriously (101). He then insults her person by sexually assaulting her. Hattie’s sexual encounter with Mr. Hyslop leads her to have a mental breakdown, though the male doctor who treats her nonsensically blames the “overstimulation caused by the reading and writing for her thesis” to be the reason why she has fallen ill (93). Silko uses these examples to confine Hattie’s education within the world of men’s rules, as she is constantly under the

influence of the male perspective. Though she becomes “educated,” in terms of patriarchal methods, her self-knowledge is completely hampered. Her attempts to express herself and her budding feminism are condemned and rebuked by the men in her life.

Silko presents Hattie’s suppressed sexual desires and discontentedness with life as another aspect of her patriarchal cage. Hattie’s sexuality is confined within the realm of high society women. Ironically, even in this all-female setting, women are still influenced by male institutions because of their own upbringing in patriarchal society and the moral standards passed down to them from their own mothers. In her youth, Hattie’s mother is adamant about Hattie attending catechism classes, where the nuns outline clearly the “sins of the flesh” (94). These classes are also the only times Hattie is able to interact with other women her age, and she finds they “talked about nothing but young men and marriage,” forcing her to follow their examples into a “traditional” woman’s life, or else be shunned (95). Hattie’s mother also instills in her daughter a fear of childbirth, which Hattie carries throughout her life. She unwillingly falls into another trap that confines her feminine spirit through her marriage to the strictly traditional Edward Palmer. He is skeptical about accepting feminist ideas (including Hattie’s) that contradict his own conservative views, and refuses to tend to his wife’s sexual needs because he himself is satisfied with a sexless marriage. Silko portrays Hattie as being powerless in defining her sexuality; her mother, her peers, and her husband all direct Hattie’s opinion on sex, which is fearful and accepting of the limited affection her husband offers her. In her essay “Landscapes of Miracles and Matriarchy in Silko’s *Gardens in the Dunes*,” Mary Magoulick explains that, “matriarchy cannot fully exist within a patriarchy” (22). Thus, Hattie’s struggling independence and feminism cannot be fully expressed within her patriarchal existence.

Silko contradicts Hattie’s unnatural repression under patriarchy with the Native American characters Sister Salt, who comes from a fictitious tribe called the Sand Lizard people, and the twins Maytha and Vedna, who work with Sister Salt washing clothes and beer mugs. The twins, in

contrast to Hattie, refuse to be trapped by anyone's guidelines for living, as can be seen by their ambitious plan to save up enough money to buy their own land. Silko informs us at the end of the novel that they "managed to save up enough money so they didn't have to live in a wallow of green beer" (Silko 475). They have livestock, a variety of plants, and are living a good life. When asked if they are "at least pregnant or engaged...they all laughed and shook their heads," a rejection of any male help that may have otherwise been seen as "necessary" to survive as a woman (475). Hattie's patriarchal education has taught her that women cannot survive on their own, and need men to give meaning to their lives. The twins clearly prove this idea wrong.

Silko also uses the Native American women show the power that results from self-determined sexuality, a power that is unavailable to Hattie. Sister Salt and the twins approach the views of sexuality much differently to Hattie. They accept sex as pleasurable experience, and relish in performing the act selfishly. Realizing they can use sex for more than just pleasure, they happily take their "free choice of the men willing to pay a dime for fun in the tall grass," and make money off of their gratification (218). This is partially where the money for the twins' land comes from. The Native American women have complete control over their sexuality, instead of others being in command of it like in Hattie's case. Though Sister Salt has a deeper connection to the man Candy than to any other man in the story, she does not feel ashamed about sleeping with Charlie, her side lover, and others. "Her body belonged to her-it was none of [Candy's] business," Silko explains within the tale (218). This free sexuality goes against all that Hattie was taught in her catechism classes, but Sister Salt uses sex as a tool, not as a shameful moral burden: "Old-time Sand Lizard people believed sex with strangers was advantageous because it created a happy atmosphere to benefit commerce and exchange with strangers," and it "was valued for alliances and friendships that might be made" (218, 202). Half-breed babies resulting from free sex were welcomed into the family. Sister Salt also uses her sexuality as a weapon against men, such as when the twins fear the wrath of a policeman. Sister Salt comforts them by saying, "Don't even worry about him [...] I know where to touch him so he won't tell anyone" (206).

Silko deliberately constructs a distance between Hattie and the Native American women, created by Hattie's difference in opinion regarding

self-sufficiency and free sexuality. Silko desires the reader to realize the existence of a "collective sisterhood" that all independent women are innately part of, if they disengage themselves from patriarchy. While the rest of the women in the novel support each other and welcome each other's company, Hattie feels completely alone because she is still under the influence of patriarchy. The deep connection of unity between Sister Salt and the twins is Silko's representation of a "collective sisterhood." With their free sexuality and control over when and which men come into their womanly sphere, these three females live alone together and take care of each other like a mother would a daughter, or like sisters. When Sister Salt is grieving for the loss of her little sister, it is the twins who hand feed her to keep her alive. Though Sister Salt is from the Sand Lizard people, a tribe normally left out of Indian pacts because of their strange practices, the twins do not ostracize her, but "[tease] her to cheer her because they [love] her" (201). When the twins plan on buying their own land, they offer to let Sister Salt live with them, even though she has almost no money to chip in for support. Nevertheless, they all do what they can to insure the collective survival of them all.

At the end of her storytelling, Silko finally provides Hattie the setting to be welcomed into this "collective sisterhood." After divorcing her husband and freeing herself from his constrictive hold, Hattie adds to the feelings of sisterhood by reuniting Sister Salt with her little sister. Initially, when Hattie first meets Sister Salt and the twins, she feels very distant from them. Hattie watches as "the girls [chatter] happily inside, laughing altogether" (409). At this time, Hattie begins feeling a "dreadful sense of how alone she [is]" (410). She decides not to return to the controlling life of her upbringing (as many people expected her to do as a divorcee), and instead stays to help Sister Salt and the twins build a home out west, beginning her journey into the atmosphere of a collective sisterhood. However, while trying to get supplies to Sister Salt, Hattie is raped, beaten, and mugged by white men who live in the area. This is because "the White people [of out west] disapprove of her as an independent woman who helps Indians" (Magoulick 26). Since Hattie has rejected returning to her former patriarchal existence, these white men who accept patriarchy see her differently, and attempt to snuff out her newly recognized feminist feelings. When Hattie is left bleeding and dying by the side of the road, it is unknown Indian women that find her. They wrap her naked body in a

blanket and “[tell] her not to be afraid” (Silko 457). Silko shows that, since Hattie has rejected her patriarchal life, the many sisters that exist all around her are there to support her. Hattie drags her sick body to stay with Sister Salt while she recovers. While there, a Ghost Dance is initiated, and Sister Salt promises to pray and dance for her. Hattie said it is “the dancers that [save] her life;” it is her collection of women friends who promise to dance for her that bring back her health (471). The “collective sisterhood” finally accepts her, and brings her body and spirit back to life.

Thus, through her representation of repressive upbringings under patriarchy, positive portrayal of self-determined sexuality, and assertion of the existence of a “collective sisterhood” outside the influence of Western culture, Silko effectively expresses her desire for women to break free from the patriarchal restraints that attempt to control them. She presents females with an alternative life, living as a part of a collective sisterhood that promises to offer them infinitely more independence and happiness than the patriarchal world is able to provide. Sister Salt and the twins find freedom when they reject the restrictions of

accepted patriarchal society, and embrace their free sexuality, dismissing the moral standards like those placed on Hattie. None of them have permanent men in their lives, but they are infinitely happier than Hattie, who struggles to figure out why she is so discontent and alone. Hattie, though, feels her desire for sex and independence growing as the novel progresses. In the end, she leaves her sexless marriage, breaks free of the control of patriarchy, and allows the safety of the “collective sisterhood” to envelop her. She sets herself free, and her sisters are there to welcome her into this new gynocentric world she never knew existed.

Works Cited

Magoulick, Mary. "Landscapes of Miracles and Matriarchy in Silko's *Gardens in the Dunes*." *Reading Leslie Marmon Silko: Critical Perspectives through Gardens in the Dunes*. Ed. Laura Colucci. Pisa: Edizioni Plus-Pisa University Press, 2007.

Silko, Leslie Marmon. *Gardens in the Dunes*. New York: Simon and Schuster Paperbacks, 2005.

A New Way For Life: Striving for Post-Humanity in Octavia Butler's Fiction

Daniel Hart

*All that you touch
You Change.*

*All that you Change
Changes you.*

*The only lasting truth
Is Change.*

*God
Is Change.
(Parable 3)*

Octavia Butler makes this powerful statement the opening to her first Earthseed novel. Much of Butler's work is centered on the theme of change. In order to better oneself and humanity, change must happen. Many gay rights theorists echo this theme of change with a less optimistic feeling: queer culture has seen much progress, but little change. Butler notes that change exists as neither better nor worse; a concept so deified as change cannot be described with a simple binary. Change is. Lilith, the protagonist of Butler's *Xenogenesis* trilogy, asks her captor to explain himself and the changes his alien race is inflicting upon humanity: ““You think destroying what was left of our cultures will make us better?”” The alien responds calmly, ““No. Only different”” (*Dawn* 35). Change and difference are vital aspects of life.

Butler's imagination and forays into the realm of science fiction allow her to explore extensively these themes of change and transcending humanity. The first two books of the *Xenogenesis* trilogy, *Dawn* and *Adulthood Rites*, follow the remnants of humanity after a devastating nuclear war. These last humans are rescued from the dying earth by an alien species called the Oankali, a race genetically obligated to gather information on other species and adapt old genetic structures into new biracial constructs with the data. Many humans resist because such an adaptation creates an entirely different species and ultimately means the end of mankind. In *Dawn*, the first book of the trilogy, the Oankali choose Lilith to lead the surviving humans. The Oankali bond with the humans to re-produce a new species—the resulting species will not be human or Oankali. Because of this radical change, Lilith struggles to personally accept the fate of

humanity while teaching the resistant humans. Akin, the protagonist of *Adulthood Rites* and Lilith's son, is a product of the Oankali's experimentation. He represents the first human-born male construct, made from the genetic structure of humans and aliens. After being kidnapped by the humans, he is abandoned by his family there so he may collect information and learn about his captors. Akin's humanity causes him to sympathize with these resistors, making him advocate to the rest of the Oankali in defense of the humans. “Bloodchild” is a short story by Octavia Butler in which a group of humans, fleeing from persecution on Earth, form a symbiotic relationship with an insect-like species on another world. The aliens use the humans as a host for their eggs while in return providing shelter, food, longevity and a new kind of family. The narrative follows Gan, a human male, and his struggle to accept his role in the family and with his alien implanter T'Gatoi.

Many of the characters in Butler's works are dehumanized, reflecting the treatment of minorities in a white patriarchal society. She even insinuates that such behavior, such cruel dehumanization, is natural to humanity. It is human nature to be an extremely intelligent people, yet still cling to hierarchical tendencies. Octavia Butler suggests that humanity needs to surpass itself, to adapt and change. Rather than re-humanize in response to dehumanization, the human race must post-humanize and transcend its natural defects. Because even ideal humans are imperfect, post-humanization is the only way to succeed and escape the inevitability of extinction.

A New Sexuality

The existence of a third sex among the Oankali is problematic for the humans who encounter them, but the overall characteristics of these unsexed “ooloi” are positive and what they provide humanity is undeniably beneficial. Ooloi “perceive things differently” than humans, typifying them as unique and fascinating (*Dawn* 22). In their drive to gather information, the ooloi partner with human couples and facilitate extraordinary sex. Lilith admits about her unsexed ooloi partner, “Nikanj could give her an intimacy with Joseph that was beyond ordinary human

experience,” even though the thought of sex with something so otherworldly repulses her (*Dawn* 168). As Nancy Jesser describes in her article “Blood, Genes and Gender in Octavia Butler’s *Kindred* and *Dawn*,” the ooloi “produce the acute sense of dissociation of self, felt by the humans as they are programmed to find pleasure in an alternative/perverse sexuality” (46). Xenophobia and homophobia (many humans perceive ooloi as masculine, even though they are ungendered) play a large role in human resistance to the ooloi.

The Oankali have unique sexual relationships and expressions of sexuality. Lilith’s ooloi, Nikanj, confuses Lilith when it uses the term “share sex” rather than “have sex” (*Dawn* 101). To the Oankali, expressions of sexuality are shared through mutual contact of sensory tentacles. Therefore, all parties experience the pleasure. Oankali form family relationships with a male, a female, and an ooloi. The male and female are related, often brother and sister or close cousins. Ooloi, the facilitators of sexual relationships, come from another family and are thought of as ‘treasured strangers.’

Furthermore, Oankali know intuitively about these relationships. Akin insists when questioned, “I’ve known since I touched my sibling... we were two-thirds of a reproductive unit... My body knows that” (*Adulthood Rites* 137). He knows this because of the Oankali in him, but after his metamorphosis, the humanity in him makes pairing with his sister feel weird and incestuous (*Adulthood Rites* 231). Tino, the new human male in Lilith’s family, to adjusts to the different sexuality with difficulty because of his human tendencies. He finds it awkward to relate to Dichaan, the Oankali male, because “he never expected to have a male mate,” and, even though their sexual expression pleases Tino, he likely maintains residual homophobia (*Adulthood Rites* 179).

Humans make gender assumptions and prescribe gender characteristics. Oankali, however, smash gender prescription and transcend traditional normativity, only adapting gender when necessary for balance. Upon first meeting one of her alien hosts Lilith asks if the alien is a male or female, to which he responds, “It’s wrong to assume that I must be a sex you are familiar with” (*Dawn*, 12). Additionally, Oankali children are sexless until adolescence at which point they take on the gender that best balances their family structure. Twice in *Dawn* men assume their ooloi are more male than neuter where Lilith “[takes] their word for what they are” (*Dawn* 92, 179). Some humans question Lilith’s femininity because of her augmented

strength, saying, “‘Only a man can fight that way’” (*Dawn* 154). Akin often notices that human men he interacts with try to intimidate him with their size; he finds this ridiculous (*Adulthood Rites*). He knows that their gendered spectacles of domination mean nothing to him as his strength significantly surpasses all of theirs.

Alien women in Butler’s work tend to be heartier, bigger and more robust than males. Oankali females are nearly twice as large as ooloi and are significantly bigger than Oankali males. Women, they explain, need to be larger in order to best nurture and protect children still in the womb. In “Bloodchild,” Tlic females live to extreme ages: “[T’Gatoi] was nearly three time my mother’s present age, yet would still be young when my mother died of age” (“Bloodchild” 8). They measure three meters long and are strong enough to overpower any human. Compared to Females, Male Tlic live much shorter lives. Butler’s reversals of physical power dynamics logically support her model for a new humanity.

The introduction of a new species into sexual relationships requires a restructuring of sexual ideas and behaviors. As Eva Federmayer writes, “The utopian thrust of Butler’s trilogy, then, crucially depends on cyborg identities and [cooperation] with aliens” (102). Expectedly, this cooperation causes backlash among many of the human characters in Butler’s novels. Joseph, Lilith’s first human mate, distrusts the experiences Nikanj gives him, insisting they are illusions (*Dawn* 198). Now that Lilith has bonded with Nikanj, her new sexual partner Tino cannot separate the two of them because “Lilith and Nikanj were a pair somehow” (*Adulthood Rites* 43). Castleberry believes the introduction of an alien species can be representative of homosexuality: “The frontrunners of using homosexuality in [science fiction] didn’t use it directly—they used intermediaries, such as aliens or a “third” sex to get their point across” (13). This is unlikely to be Butler’s intention because Oankali pair solely for reproduction. However, Butler may be suggesting a naturalization of alternative sexualities because the Oankali can communicate—as in talk, get feelings across, and provide immense pleasure—with all other Oankali people. Family joining for them does require polyamory, but no limitations exist on pleasurable interactions outside the family unit with people of any gender.

A New Family Structure

Lilith's characterization in the *Xenogenesis* novels shapes the nature and the tone of the novels, foreshadowing the role the Oankali have for her. *Dawn* opens with a section entitled "Womb" and Lilith curled in a fetal position, setting the tone of motherhood and deliverance for the trilogy. Also, before the nuclear war on earth, Lilith birthed her children naturally and not in a hospital—she didn't "like the idea of being treated as though [she] were sick" (*Dawn* 94). Because she was chosen by the aliens, Lilith, "has to negotiate her role as 'race mother'" (Federmayer 102). The Oankali find her strength, adaptability and intelligence suit Lilith perfectly for her task of assimilating the first humans.

Outside of the novels, the name Lilith bears a great deal of significance when analyzed concurrently with the protagonist's role of 'race mother' in the novels. Indeed, Lilith admits it is an "unusual name loaded with bad connotations." (*Adulthood Rites* 36). Lilith, a prominent figure in Judeo-Christian mythology, takes on numerous incarnations throughout Europe and the Middle East (Dame, Rivlin, Wenkart 7). She is said to be the first woman God created, formed "just as he had formed Adam, from pure dust" (5). Unlike Eve, Lilith denies Adam's claim to dominance. When Adam insisted she lie beneath him during sex and force her obedience, "Lilith chose to leave Adam and live alone by the Red Sea" (5). Here she experienced great sexual freedom, copulating with half-humans and birthing demonic babies. Many incarnations of the myth portray her as a demon or a witch, at fault for such horrific deeds as infant mortality and nocturnal emissions. But feminist religious scholars have since deconstructed the myth of Lilith and find her actions empowering. In fact, "Some scholars suggest that the extant myths may have been rewritten from an earlier gynocentric mythology." Perhaps "the male mind, in order to cope with sharp-witted females, could only imagine a demonic explanation" (10). Butler's choice to name the protagonist Lilith underlines her role in the new civilization. Furthermore, the Oankali place Lilith and her people in a pollutant-free and genetically supplemented jungle wilderness, bearing uncanny similarity to the Garden of Eden. Maybe Butler wishes to indicate that humankind would have been different and better had the biblical Lilith been the woman Adam paired with and that this nuclear war is a chance to start that superior humanity.

Though much about them is different, the Oankali share many similarities to the positive characteristics of humans. Their definition of "parenting" bears similarity to a human's: "To teach, to give comfort, to feed and clothe, to guide them through and interpret what will be, for them, a new and frightening world" (*Dawn* 116). An important note is the gender neutrality of the word "parent," as the alien family structure does not place emphasis on either "fathering" or "mothering." Oankali still experience feelings much as humans do. They require balanced nourishment and exercise. And they have an easy time relating to the humans on personal levels. Butler's characterization of the Oankali illustrates as much resemblance to humans as possible while still keeping a degree of foreignness. By doing this she paints them as an exemplary species, a level of being humans could possess but for the disposition favorable to power relationships.

Butler's aliens have unique family structures. The Oankali's involves up to five parents and the families are perfectly balanced. "Akin spent some part of the day with each of his parents" rather than being unequally raised solely by one of his parents (*Adulthood Rites* 12). Akin recognizes this need for balance when he admits already thinking of his unborn, ungendered sibling as a sister (13). Dichaan, the Oankali male involved in Lilith's family unit, feels the new human male, Tino, in the group provides "a balance found after painful years of imbalance." He feels the imbalance more than any other because he "had been born to work with a Human male parallel" (64). In this family there is no concept of adultery because each Oankali can communicate feelings, sensations and pleasures to any other being; family relationships are to benefit child rearing and have no undertones of possession (224). The Tlic family structure is well controlled: only one female allowed to breed. Other sisters remain sterile and provide unfertilized eggs to heighten the health of the humans. Males seldom factor into this equation at all; they live such brief lives and spend as much of it as possible copulating. But the Tlic incorporate humans into their family in order to best suit both species' needs, causing everyone to have "a personal stake in keeping the peace" ("Bloodchild" 12). Gan's mother and T'Gatoi both went through adolescence together and T'Gatoi is equally as protective of Gan as his mother. When another Tlic learns her surrogate, Lomas, has just violently had the implanted larva removed, the first thing she asks is whether Lomas is safe ("Bloodchild" 18). These

family structures indicate a multitude of ways to organize families in order that all parties involved are safe, satisfied and well taken care of.

Butler's fiction imagines a great deal of changes in birthing routines and standards. After publishing these books under the name *Xenogenesis*, Butler made the decision to change the name to *Lilith's Brood*, a more gripping and evocatively natal title. When Joseph dies and Lilith and Nikanj are left without a male to balance the relationship, Nikanj impregnates Lilith with Joseph's child by simply manipulating her cell structure and using genetic remnants of Joseph's biology (261). This child, immaculately conceived, will be the first genetic mix of both species. Already discussed is the importance of Lilith's name, both in relation to mythology and the Genesis of the human race; now that the protagonist Lilith will be bearing myriad mongrel children the comparison is even more befitting. Construct children, as Akin describes, can remember their time in the womb vividly (*Adulthood Rites* 3). During the Oankali birth, "[The Oankali female] needed all her mates near her, touching her, needed to be able to link into them and feel the parts of her child that had come from them" (*Adulthood Rites* 82). This need for closeness parallels human birth, but the difference is the other Oankali's recognition of their absolute need to be present. The text suggests that humans should adapt to value this need as well.

The aliens impose a few birthing restrictions in both the novels and the short story. After pairing with the ooloi, the humans no longer have a desire to mate with each other and are actually repulsed when an ooloi is not present for the interaction (*Dawn* 234). Nikanj admits the Oankali did this to curb dangerous over-population by rebellious humans (*Adulthood Rites* 42). In "Bloodchild," designated family members provide sterile eggs to give health benefits and pleasure to humans. The suppliers of these eggs are the "sterile sisters" because there is only "one fertile female in every lot" ("Bloodchild" 23). Additionally, Tlic officials safeguard and carefully dole out humans to deserving Tlic families. T'Gatoi recollects a time when humans were caged and treated as nothing more than animals. With the current arrangement, T'Gatoi manages the Preserve where the humans stay and protects them from greedy and desperate Tlic would-be rapists.

Not only do birthing rituals change in Butler's work, in "Bloodchild" she reverses the roles of birth to balance the gendered roles society dictates.

An Afterword to "Bloodchild" declares Butler wanted to "write a story in which a man chose to become pregnant *not* through some sort of misplaced competitiveness to prove that a man could do anything a woman could do, not because he was forced to, not even out of curiosity" ("Bloodchild" 30). As Elyce Helford points out, the world of the Tlic is one of "a biologically determined matriarchy" (10). The females hold the positions of power because such comes naturally. In addition, when T'Gatoi goes to impregnate Gan with her eggs, she penetrates him to implant her fertilized egg, providing a description that "clearly complicates the traditional gendering of sexual imagery" (Helford 9). This birthing process bases itself on alien biology and its impact on the human body, just as pre-feminist conceptions of intercourse were centered on a male (subject) affecting the female (object). Displacing the act of penetration onto males deconstructs typical sexual narratives.

A New Humanity

Butler uses her fiction to point out the many shortcomings and deficiencies of humanity. Before the novel begins, when the Oankali first encountered humans, "It hurt some of them to touch" these humans; they "had never before seen so much life and so much death in one being" (*Dawn* 25). This "the human contradiction,"—as it is referred to in the rest of the book—defines humans as intelligent but also possessing a propensity toward hierarchical behavior. The latter genetic tendency selects in breeding and holds humankind back, limiting it. Many of the humans Lilith teaches feel that pairing off and mating with each other is necessary for survival, to the point that they would rape in order to accomplish this (*Dawn* 187). To Lilith this is unacceptable: "'There'll be no rape here... Nobody here is property. Nobody here has the right to the use of anybody else's body. There'll be no back-to-the-Stone-Age, caveman bullshit!'" (*Dawn* 188). That these humans believe rape the only option for human survival, and that Butler uses Lilith's dialogue to hearken back to humanity's conception, indicates a flawed human mentality. Human males cannot feel the magic of bearing a child inside of them as females can. However, the Oankali can simulate that feeling and share it between the genders, allowing for greater understanding and empathy (*Adulthood Rites* 55). When Akin returns to the village in which he spent much of his

childhood, he encounters a mother-like figure, the one who nurtured him while he was stranded with the humans, in poor physical health. He asks to help her by entering her nervous system and fixing the problems directly. But her human husband, suspicious and petty, does not trust Akin to heal his wife (*Adulthood Rites* 248). Another representation of humanity's appalling nature, this human would put his hatred and racist feelings before his wife's physical wellbeing.

Despite human deficiencies, they are still undeniably alluring to the Oankali and beneficial to the Tlic. Humans inexorably draw the Oankali because they present such a curious and strong genetic contradiction. Also, during the gene trade the humans gave the Oankali cancer, a fascinating anomaly in the eyes of the new species. Ooloi have the ability to manipulate cellular biology and structure; with the existence of cancer they see endless possibilities like limb regrowth, enhanced musculature and beyond. Humans in Butler's fiction often augment alien lifestyles. After human contact, the previously stagnating Tlic begin to flourish. T'Gatoi says to Gan: "'because your people arrived, we are relearning what it means to be a healthy, thriving people'" ("Bloodchild" 25). She suggests that neither species was perfect before but now that they have each other they can learn how to cooperate and live better. It seems Butler is not completely skeptical of human capability.

Because humans fear change, they come into direct conflict with the Oankali, a species with a genetic imperative toward change. When presented with an offer to be changed by an ooloi, Lilith reacts negatively even though the change would give her increased strength, resistance to disease and an eidetic memory. Peter, one of the most resistant and hateful humans Lilith has to teach, finds the manipulation by the ooloi frightening and emasculating (*Dawn* 203). Butler takes this opportunity to express how removing dynamics of masculinity from human nature is ultimately positive. While captive, Akin constantly questions the human motives to remain unhealthy and infertile when pairing with an ooloi and working logically with the Oankali solves these problems (*Adulthood Rites* 69). Conversely, Tino, a full human, laments leaving his family and joining with the Oankali because he feels he is "a traitor to [his] people" (*Adulthood Rites* 180). This view comes from him being one of the last humans who still lays claim to complete humanity yet he helps the other species proliferate half-breeds. However, Nikanj makes an interesting parallel:

'Examine Tino. Inside him, so many different things are working together to keep him alive. Inside his cells, mitochondria, a previously independent form of life, have found a haven and trade their ability to synthesize proteins and metabolize fats for room to live and reproduce. We're in his cells too now, and the cells have accepted us. One Oankali organism within each cell, dividing with each cell, extending life, and resisting disease... They could not exist without symbiotic relationships with other creatures. Yet such relationships frighten them.' (*Adulthood Rites* 183)

Nikanj's also cites the quantities of bacteria that also find shelter in the human body. Its unflappable logic still conflicts with human fear and irrationality. The humans in these novels do not realize humanity and Earth, as they stand after the nuclear war, are doomed. Each Oankali comprehends that coming to the aid of humans was the only viable option.

What the Oankali and the Tlic do to humans can be viewed as colonization or slavery, but such an analysis is too simplistic; Butler provides numerous cues that speak to a different, new kind of relationship. Holden writes, "The Oankali consumption of difference turns out to be not much better than the human repudiation of difference; both result in a domination and/or erasure of the 'other'" (51). This is not true. Each of the constructs has both Oankali *and* human genes equally present in their biology. Rather than an erasure, the Oankali provide a single, more perfect species with the union of both. Also, the Oankali split up into three groups: Dins to bond with the new species on the planet; Toaht to bond with the new species in the safety of their mother ship; and Akjai to remain as they were before bonding in case the mix turns out disastrous. After his time with the humans, Akin understands their rationale and realizes that the humans need an Akjai group in order to have another chance at life their own way. Though many Oankali resist, Akin convinces them and the dissident, pure humans are given a second home on Mars. Additionally, when Akin communes with one of the animals the Oankali use for transportation he makes an interesting discovery: "Oankali simply bred animals who did not want to stray and who enjoyed doing what they were intended to do. They were also pleased to be rewarded with new sensations or pleasurable familiar sensations" (*Adulthood Rites* 203). Because the animal maintains a desire to remain in service, the best relationship the two species can

have is a symbiosis. This relationship mirrors White Man's Burden to a degree, but with crucial differences. First, the Oankali can directly link into the nervous system of other animals and people to feel with them. Any dissatisfaction, pain, or injustice would only be magnified in such a union. Therefore, all of the creatures the Oankali meet and bond with benefit from symbiotic rather than colonial relationships. T'Gatoi, and the other Tlic government officials, "Parceled [humans] out to the desperate and sold [them] to the rich and powerful for their political support" ("Bloodchild" 5). But Gan makes clear that the humans are an independent people. Before T'Gatoi advocated for humans, the Tlic would break up human families and use their bodies without any regard. Current relations changed for the betterment of humans. Without the protection of T'Gatoi's political faction, the hoards of desperate Tlic would enslave the humans rather than cohabit and respect them as a sentient species. T'Gatoi assures Gan he, rather than just the role he plays, bears importance to her: "You know you aren't animals to us" ("Bloodchild" 24). Relating Tlic/human to human/animal relationships alludes in an abstract way to marginalization by patriarchal society and how it is similar to human subjection of animals. Additionally, Gan demands he keep the gun that is illegal in order to balance the relationship between them, insisting, "If we're not your animals... accept the risk" ("Bloodchild" 26). Gan's actions equalize the two characters, showing a complex series of interactions that blur the lines of subject/object. Though analyzing Butler's work in terms of colonizer/colonized relationships is appealing and thought provoking, the argument carries less weight than some commentators think. These new relationships must be thought of as completely separate from the dynamics of modern human relationships because Butler constructs a new kind of life, a new humanity.

Conclusion

Critics discuss Butler's work as not post-gender because many of her main characters reinforce traditional female roles. According to Nancy Jesser, "Butler offers some counter narratives to the most retrograde of evolutionary biology, [but] she continues to be very much bound to the explanatory and predictive power of genetic sexual difference" (42). This interpretation, however, is insufficient. Butler may use traditional sex roles to describe her humans and her alien

reproduction, but she only does so to recognize that gender differences do indeed exist today. Her exemplary species, the Oankali, use gender for the mating and parenting necessities, showing that since sex remains reproductively essential, it should be *socially imposed* gender differences that go by the wayside. While the deficient humans maintain these differences and the male tendency to dominate, the male and female Oankali have few differences and both recognize their purpose and requirement in social life.

Butler's work pleads for humankind to fix itself and remedy its shortcomings; she appeals humans to embrace change. In an interview with Charlie Rose on an NPR broadcast Butler discusses these failings: "There are times I know we can do better than we have because we have done better—but I don't know if we will. I know that I want us to do better, I want us to focus on other things" (2000). In the fantastic worlds Butler creates humanity can improve by casting off the defects and imperfections of humanness. Her characters, dehumanized in some way or form, find their solution is not to re-humanize, but rather to transcend character weaknesses inherent with "the human contradiction." Is it possible for modern humanity to accomplish such a goal? Butler, rather than constructing a post-gendered world, creates a species that balances the roles of different genders while still recognizing the reality of each sex. Even with Butler's appeals, though, it is unlikely mankind will go off the path of expediency and desert hierarchic predilections—barring, perhaps, nuclear war.

Works Cited

- Butler, Octavia E. *Adulthood Rites*. New York: Warner Books, Inc., 1988.
---, *Bloodchild and Other Stories*. New York: Four Walls Eight Windows, 1995.
---, *Dawn*. New York: Warner Books, Inc., 1987.
---, *Parable of the Sower*. New York: Warner Books, Inc., 1993.

Helford, Elyce Rae. "'Would you really rather die than bear my young?': The Construction of Gender, Race and Species in Octavia E. Butler's 'Bloodchild.'" *African American Review*. Summer 1994: 259-272. EBSCOhost. Milne Lib., Geneseo, NY.
<http://search.serialsolutions.com>

Charlie Rose: Live at NPR. Natl. Public Radio. 1
June 2000.
<<http://www.youtube.com/watch?v=66pu-Miq4tk>>

Dame, Enid, Lilly Rivlin and Henny Wenkart ed.
Which Lilith? Feminist Writers Re-Create the World's First Woman. Northvale: Jason Aronson, Inc., 1998.

Holden, Rebecca J. "The High Costs of Cyborg Survival: Octavia Butler's *Xenogenesis* Trilogy." *Foundation* 72. Spring 1998: 49-57.

Jesser, Nancy. "Blood, Genes and Gender in Octavia Butler's *Kindred* and *Dawn*. *Extrapolation*. Spring 2002: 36-61.

Castleberry, Traci N. "Twisting the Other: Using a 'Third' Sex to Represent Homosexuality in Science Fiction." *The New York Review of Science Fiction*. January 2009: 13-17.

Federmayer, Eva. "Octavia Butler's Maternal Cyborgs: The Black Female World of the *Xenogenesis* Trilogy." *Anatomy of Science Fiction*. 2006: 95-108.

Faith and Authority in Euripides' Medea and The Bible

Caitlin Kowalewski

As people who are essentially foreigners, whether geographically or ideologically, Medea, Jesus, and his Apostles are forced into positions of subservience by the societies in which they live. They all exist as minorities, whose actions conflict with social norms. Because of the hostility they receive from figures of authority trying to preserve these norms, their relationships with even *higher* authorities, gods, are highly emphasized. However, the perceptions these worshippers have of their gods have a great influence on how they relate to and trust them, and as a result, impact how they clash with figures of power on earth.

In Euripides' *Medea*, Medea struggles with her own subjugation as a foreigner, a woman, and an image of daunting mysticism. She is a fiercely autonomous woman –a paradoxical description in ancient Greece—and additionally reputed as a murderer and enchantress; thus, she is a fairly unorthodox and isolated figure in the nation of Corinth.

As an outsider in a culture that condemns her as a criminal, and whose actions are considered sinful, it is difficult for Medea to trust that the gods will take any active role in her quest for revenge against her husband, who has abandoned her. Though a descendant of Apollo, she does not have his constant presence (like Aeneas does Venus) to assure that her wishes will be fulfilled. For this reason, Medea navigates her life independently and in a much more manipulative, subtle manner than a person whose decisions are divinely affirmed. She says to Creon, the Corinthian king who has banished her out of fear, “Don’t let *me* alarm you, Creon. I’m in no position- / A woman- to wrong a king. You have done me no wrong... So now I bear no grudge against / *Your* happiness: marry your daughter to him [Jason, her husband], and good luck / To you both” (Euripides 26). From Medea’s previous lamentations we know that this is the absolute opposite of what she wishes, but Medea realizes that she is isolated amongst people as a foreigner, and unsure of her favor with her gods as a criminal. Accordingly, she can trust only her own abilities, and manipulates her image to appease those to whom she must submit.

Medea’s relationship with her mortal and divine authorities develops quite interestingly throughout the play, in that she holds her human

counterparts accountable to the higher law of the gods who have potentially abandoned her. While she may remain unsure about her own goodness, she is confident in the fact that her enemies have wronged her, and will be judged by a higher power for doing so. In her interactions with Creon, we can see how this respect for the actions of gods results in disdain for those of men. She says, “Oh, what a fool! / By banishing me at once he [Creon] could have thwarted me / Utterly; instead, he allows me to remain one day” (28). This contempt for the power and decisions of men is revealed in her condemnation of Jason as well. She bitterly convicts him through the eyes of a higher power: “Zeus the father of all / Knows well what service I once rendered you, and how / You have repaid me. You were mistaken if you thought/ You could dishonor my bed and live a pleasant life / And laugh at me” (59). Her disrespect is scathing when she reveals it, and displays her invocation of divine justice, but even more impressive is the manner in which she conceals this contempt to get what she wants.

When appealing to Creon to remain a day in Corinth, she intelligently engages his sympathies as a father, cleverly taking her image into her own hands and assuring him that “I know what trouble is; I have no need of more” (27). Yet after he leaves, it is clear that Medea submits to no unreliable authority, and acts only upon her own. She says, “Do you think I would have fawned so on this man, / Except to gain my purpose, carry out my schemes?... Today three of my enemies I shall strike dead” (28). Medea cannot rely on Creon for any kind of justice or sympathy with her plight, nor can she trust that Fate will not banish her “without resource” (29). Unaided by her gods, she must appease human authority and depend on herself for the execution of her will.

When dealing with friendly powers, her manipulation of authority figures holds no exception. Upon meeting with Aegeus, king of Athens, she seizes her opportunity for safety by promising him what he most desires: “Receive me in Athens; give me a welcome on your house... I know certain drugs / Whose power will put an end to your sterility. / I promise you shall beget children” (39). Though she does not respect his power as a mortal ruler, Medea successfully

presents an image of humility and piteousness to ensure her own safety. She does not demand or wait to be offered sanctuary in his kingdom, faithfully believing the gods will arrange for her salvation, but rather bargains with a leader whose desires she can manipulate.

The manner in which Medea interacts with mortal figures of authority is quite different from that of Jesus and his Apostles in *The Bible*, based on the fact that their relationship with *divine* authority is much more tangible, and for them, reliable. The Apostles distinguish their faith from that of Medea in that they believe God will stand by them unequivocally if they follow His law. Given a clear set of instructions in the Ten Commandments and Jesus, who performs miracles before their very eyes, the Apostles believe in a god whose word and works surround them, and whose instructions will presumably secure their acceptance into the kingdom of Heaven. For the Apostles and Jesus, mortal law is feeble in the face of God's divine command, and if the former conflicts with the latter, they believe it should simply be abandoned.

Because of this belief, Jesus and the Apostles frequently find themselves at odds with human leaders and their laws, a fact clearly seen in the Book of Luke and the Acts of the Apostles. Because of their devout belief that God will guide the righteous, they are not hesitant to advocate what they believe, or to disobey unjust laws. They treat rulers and kings the same way they treat the common worker, preaching that they should repent their sins and follow the word of God. In Luke, Jesus denounces the practices of Pharisees and lawyers, cursing their impious ways: "Woe to you Pharisees! for you love the best seat in the synagogues and salutations in the market places... Woe to you lawyers also! for you load men with burdens hard to bear, and you yourselves do not touch the burdens with one of your fingers" (Luke. 11.42-45). They do not reserve distinction or reverence for men who will hold no power in heaven.

Jesus and the Apostles not only treat authoritative figures like everyone else, but preach that they should impose their measures of equalization upon themselves. When a ruler asks him how he may attain eternal life, Jesus tells him to obey the commandments. When the ruler says he already observes them, Jesus says, " ‘One thing you still lack. Sell all that you have and distribute to the poor, and you will have treasure in heaven; and come, follow me’ " (Luke. 18.22). Jesus believes that God's is the only authority that matters, and

speaks his beliefs without hesitation, unlike Medea, because he believes that all men will be judged by God in the way He describes. There is no mystery in the teachings of Jesus' God. The Ten Commandments and the stories of Genesis and Exodus provide a clear image of God's desires and examples of how His laws should be executed. Because of this, Jesus and the Apostles have a concrete source of reference to the word of God and proof of God's covenant: his promise to protect and provide for those who follow Him. Compared to Medea, who has been shunned as a criminal and whose gods are notoriously self-interested and vindictive, it is no wonder that Medea deals more hesitantly with leaders who oppose her. She does not have the support of a sovereign God, whose will is elucidated before her and whose idea of justice she knows she can execute. She can only hope that she incurs favor in her gods' eyes, and carry out her plot for revenge on her own, without any divine protection.

In the course of the story of Jesus and his Apostles, some of the most resonant events occur in Luke and Acts, when the disciples' belief in their God's divine protection is tested by the threat of mortal violence. The conflict between physical and spiritual punishment, the immediate and the obscure, is one that the Apostles face again and again, but ultimately their unrelenting faith in God's benevolence resolves any tension or hesitation they may feel. They know that their suffering, whether it be arrest or crucifixion, will be rewarded in the afterlife as a sign of loyalty to God. In Acts, the prophet Stephen is stoned to death for condemning non-Christian authorities as being a "stiff-necked people" who resisted the Holy Spirit, and who "received the law as delivered by angels and did not keep it." When the people seize him in anger, he looks up to heaven and says, "Behold, I see the heavens opened, and the Son of man standing at the right hand of God" (Acts. 7.51). Stephen's faith in God contributes to his boldness in the face of those who despise him, and as a result, his relationship with them is one of unfeigned disregard. For Jesus and the Apostles, the societal authority only matters in service of divine authority, and is thus superseded by the rules of a benevolent and reliable God.

In both *Medea* and *The Bible*, the conflict between divine and mortal authority is one that shapes the very actions of each story's people, revealing how strongly their faith or their doubts control their opinions. Medea, a social outcast, believes that her gruesome revenge will be justified

in the eyes of her gods, but because she has no proof that these gods will vindicate her against her enemies, must associate with them in a manipulative manner, cautiously presenting herself as an object of pity and harmlessness.

Contrastingly, Jesus and his Apostles hold devout faith in the god whose laws are presented before them, and unflinchingly criticize human authorities whose powers pale before the law of an all-powerful divinity. So while all of these religious people are forced to face hostile figures of human judgment, the manner in which they do so depends greatly on the nature of their faith in divine judgment, and in their own self-perception as administrators of their will.

Works Cited

Euripides. "Medea." *Medea and Other Plays*. Trans. Philip Vellacott. London, England: Penguin Classics, 1963. 17-62. Print.

The Holy Bible: Revised Standard Version. New York: Meridian, 1974. Print.

The Second Shepherds' Play: "Coll's Speech"

Sally Profeta

Medieval mystery plays were significantly influential to English theater throughout the sixteenth and into the seventeenth century. They were well acclaimed by all facets of society not only for their appeal to the ‘religious orientation’ (Roney 1) of their audiences but also the social controversy of the time period (Greenblatt & Abrams 407). *The Second Shepherds' Play (Secunda Pastorum)* is an example from the “Towneley Cycle” (Edminster 1) or Wakefield Cycle of mystery plays that achieved great success. This play was likely an “alternative” (Roney 1) to *The First Shepherds' Play (Prima Pastorum)*, one of two Nativity plays written by an anonymous playwright from Wakefield, Yorkshire, who is widely acknowledged as the Wakefield Master. According to scholars, the Wakefield Master was most likely a “highly educated cleric” whose greatest talent was coalescing religious allusions with comedic nuances. In the *Secunda Pastorum*, the Wakefield Master’s “development of plot and character” is especially unparalleled by other mystery playwrights of his time (Greenblatt & Abrams 408).

Even though this play was meant to be a parody of the Nativity story, it also contains significant medieval socio-political commentary that is especially evident in the first scene (Edminster 1). Wakefield introduces Coll, a shepherd and tenant farmer, through a lengthy poetic soliloquy that describes Coll’s dissatisfaction and distress concerning his socio-economic status. This lament is an example of Wakefield’s use of Estate’s satire not only to present the shepherd as a singular character, but as a representative of an entire class of disgruntled, lower class citizens. He starts by describing his complaints with the weather and his physical condition, then goes on to describe how injustice and poverty is inflicted on his social class through taxation and forced obedience to their superiors, concluding with a reversion back to his normal routine (Greenblatt & Abrams 409).

In the 1994 edition of the Early English Text Society by A.C. Cawley and Martin Stevens, there are thirteen lines per stanza, (“thirteeners”), whereas they were “traditionally” printed as 9 lines per stanza with an introductory quatrain which consisted of four longer lines. Additionally, the language has been “normalized” to modern English

where applicable. The play’s rhyme scheme, according to this particular edition, is noted as a b a b a b c d d d c (Greenblatt & Abrams 408).

Firstly, one must examine the themes that are consistent throughout the soliloquy. Wakefield expresses two major themes in Coll’s speech, namely, hardship and class antagonism. Hardship is conveyed throughout Coll’s speech by way of his continuous complaints concerning melancholy, resentment, and anguish that are usually stressed at climactic moments in each stanza. For example, in the very beginning he explains his dread for winter when he describes how the “weathers are cold”, he’s “nearhand dold” or nearly “numb”, his “fingers are chapped” or roughened, and he’s all “lapped in sorrow” (“wrapped” in grief). He sympathizes with those that “never rest / Midday nor morrow”, which expresses his feeling of exhaustion. As he continues, he supplements his complaints with words such as “rammed” which can be translated as “beaten down” and statements such as “Thus live we in pain,/ Anger, and woe, By night and by day” emphasizing the constant burden of his dissatisfaction that causes him great torment (Greenblatt & Abrams 408-409).

It is imperative, however, that the audience understands the social implications of being a lower class shepherd and a farmer who cultivates the land of wealthy landowners. Thus, Wakefield uses the theme of class antagonism coupled with hardship to portray his underprivileged social status and illustrate a popular criticism on hierarchical society. This example of estate’s rivalry portrays the conflict between the three estates (clergy, aristocrats, and peasants) in medieval society. Coll describes his occupation when he states:

But we sely husbands
That walks on the moor,
For the tilth of our lands
Lies fallow as the floor.

(Greenblatt & Abrams 409)

In this line he describes himself as a “[husband]” or a “[plowman]” who originally cultivated his superior’s farmland but now does not because these landowners found it more profitable to let the soil grow fallow and create an enclosure for sheep pasture. The conversion of these peasant plowmen into shepherds has lowered their “economic worth” and thus reduced their capability to sustain their

already difficult lifestyle. Coll complains about the “abuses and predations of the upper classes” and the obvious injustice inflicted upon the peasantry by denying their right to cultivate the land (Edminster 3).

To illustrate the effects of this exploitation of the lower classes, Coll exhibits discontent by saying “In faith we are nearhands / Out of the door” meaning that in reality these husbands, including himself, are practically homeless. He describes his superiors as “these men who are lord-fest” or men who work hand in hand with feudal oppressors, who “cause the plow tarry” or inhibit efficient farming. By saying that these men “reave us our rest” Coll is depicting them as heartless oppressors that overwork their peasantry for self-gain “In point to miscarry” or to the point of absolute ruin on the part of the impoverished. With this strong emotionally driven expression of frustration, the rivalry between the upper and lower classes are apparent (Greenblatt & Abrams 409).

Furthermore, Coll gives the audience a glimpse of these lord-fest men’s self-righteousness by saying:

For may he get a paint-sleeve
Or a brooch nowadays,
Woe is him that him grieve
Or once again-says.

(Greenblatt & Abrams 410)

This suggests that these men become full of pomp and inflated pride once they are granted a position of authority by their lords, instantly becoming haughty and tyrannical towards opposition. Coll’s perspective on his oppressors is especially lucid when he states, “I were better be hanged / Than once say him nay”, which means that these men would rather have him killed than hear rebuttal (Greenblatt and Abrams 410). The sarcastic tone that Coll uses helps amplify his dissatisfaction and thus makes it easier for the audience to relate to his anger. His tone is forceful at times, for example when he says “Thus hold they us under, / Thus they bring us in blunder, / It were a great wonder And ever should we thrive,” which emphasizes the oppressive aspects of these lords rule over the husbandry and the lack of opportunity these peasants have now that they are so heavily subjugated (Greenblatt and Abrams 409). Coll describes the hierarchy more clearly when he implies that even these men who demonstrate “boast and bragance” are themselves supervised by even higher standing upper class men, or “Men that are greater” (Greenblatt & Abrams 410).

To better call attention to the emotional aspect of Coll’s situation, the Wakefield Master uses rhyme and meter to his advantage in order to achieve a desired effect on his audience. His use of rhyme is essential to epitomizing Coll’s tone from stanza to stanza. The alliterative effect he attains when he says “We are so hammed, / fortaxed, and rammed” with the use of the words “hammed” and “rammed” expresses his absolute exasperation as a result of these oppressive measures their superiors undertake. The audience can tell that the stress is placed on all three of those expressive words (Greenblatt & Abrams 409).

Furthermore, alliteration utilized through the words “napped”, “chapped”, and “lapped”, express fatigue, feeling of being worn out, and being overwhelmed respectively. All of these rhyming words by definition emphasize weariness and melancholy and are continuously stressed throughout the stanza, giving the audience a constant reminder of Coll’s emotional state and sorrowful tone. In the same way, the words “wary” and “tarry” when Coll exclaims, “Our Lady them wary! / These men that are lord-fest, / They cause the plow tarry”, both have similar effects on the audience since by definition, wary is a curse and tarry is an unwelcomed interference. Because they rhyme, the audience understands the frustration that Coll is experiencing with these gentry-men’s ill-governance (Greenblatt & Abrams 409). Additionally, one can see how Coll’s tone changes in the last stanza by looking at the rhyming words, “walk”, “talk”, and “stalk” that express a sort of monotony and passivity that Coll feels as he realizes that after sufficiently articulating his dissatisfaction, he must to his work (stalking after sheep) (Greenblatt & Abrams 410).

Throughout Coll’s lament, the Wakefield Master uses key literary techniques to demonstrate the social and political aspects of society in this time period as expressed through a character that experiences these truths firsthand. The recurring themes of hardship and class antagonism throughout the soliloquy help convey the discontent of the peasant class. The melancholy and torment the shepherds suffer is well highlighted through Wakefield’s expert use of rhyme and alliteration to reveal the suffering the peasant class must endure as a result of hierarchical tyranny. The Wakefield Master’s skillful use of these elements makes him a great influence for the future of professional English theater, and an excellent example of a great medieval mystery playwright

Works Cited

- Edminster, Warren. "Punning and Political Parody in the Second Shepherds' Play." *English Language Notes* 40.4 (2003): 1- 10.
- Greenblatt, Stephen, and M. H. Abrams. "Mystery Plays." *The Norton Anthology of English Literature*. New York: W.W. Norton, 2006. 406- 410.
- Roney, L. (1983). The Wakefield First and Second Shepherds Plays as Complements in Psychology and Parody. *Speculum: A Journal of Medieval Studies*, 58(3), 696-723.

The Richest Are Not Always the Wealthiest

Daniel Schlant

In his classic treatise on the ideals of philosophy, Boethius discusses several principles such as fortune, fame, power, and honor. Much of his commentary on these issues, which he recorded in the sixth century B.C., still rings true in today's society. Written around the year 524 while imprisoned for treason, *The Consolation of Philosophy* reflects an intersection between the classical Greek views on philosophy and relatively recently formed Christian beliefs. As a result, Boethius wrote of classic ideals, such as happiness and fortune, as well as new ideas, such as the Christian God. Despite the span of the centuries since the writing of this piece, many of the stances Boethius proposes concerning these subjects can be applied to today's world. One such case is the generosity of Allen and Violet Large, lottery winners who donated all of their winnings. The Larges' use of their draw embodies what Boethius feels is the nature of many of the concepts he analyzes in his thesis, particularly fortune, wealth, and happiness.

An article attributed to the *Associated Press* that was carried by *The Miami Herald* describes an instance in which the principles outlined by Boethius are displayed in the modern world. The article concerns an elderly couple, Allen and Violet Large, and their interesting response to winning the lottery. The Larges, who live in a rural area of Nova Scotia, won approximately \$10.9 million dollars in a July 2010 lottery draw. Upon receiving their winnings in the fall of the same year, after taking care of a few minor affairs for themselves and their immediate family, the couple promptly gave away all of the remaining cash. Ninety-eight percent of the Larges' cash prizes went to various organizations in the Lower Truro area, their hometown. Among these organizations were the local Red Cross, churches, fire departments, and cemeteries. When asked why they were so willing to donate all of the money, an action few in today's society can honestly say they would do if in the same position, Allen Large simply responded, "We have no plans. We're not travelers. We're not night-prowlers. We're not bar-hoppers" (CBS/AP). The Larges have found themselves content with the simple life, and their astounding generosity reflects this happiness. Perhaps the most heart-warming aspects of the story can be found in

one of the organizations a large portion of their money went to. A large sum of money was donated to the facilities at the hospital where Violet Large is currently undergoing treatment for cancer. Despite all of the adversity the Larges have undergone, which would leave many bitter, they remain extremely altruistic people. The fame the Larges have experienced of late is a kind of fame that is all too rare in the modern society. Scandals and crime fill the papers, and it seems that fame has an inherent notoriety. However, the Larges justly deserved fame has been thrust upon them due to their inspiring generosity, and they are treated with the highest regard as a result. Many of the ideals that Boethius describes in great detail in his work on philosophy can be applied to and viewed in the context of the Larges' story.

The Larges' incredible generosity and subsequent fame strongly supports the philosophical ideals that Boethius outlines in *The Consolation of Philosophy*. One such ideal that Boethius discusses is the concept of fortune and luck. Boethius held that one should not put too much stock in the hands that fortune has dealt them. Many curse their bad fortune, and find themselves fed up and very bitter. Others pine endlessly for the opportunity to reap the benefits that come with good fortune. Boethius, however, finds the people who are guilty of such states of mind to be foolish. Boethius put very little value in the prizes fortune awards, and the losses that one forfeits when misfortune is experienced. As he writes, while speculating on the true nature of fortune, "You will realize that you neither had, nor have lost, anything of worth through your association with her" (19). Boethius goes on to claim that most people are actually ignorant to the good blessings that fortune actually bestows on them. The personification of philosophy, a character Boethius uses in his work to explain some of the concepts he discusses, feels the need to remind Boethius of the blessings he has been granted. Boethius is upset that he can no longer enjoy the freedoms he once did, but Philosophy chides him. She reminds him that some inhabitants of the world would be grateful to be in Boethius' position. He has a wife and children who care about him, which should be one of the main sources of meaning and happiness for Boethius. Boethius holds that most men largely place value

and their estimation of their fortune in worldly possessions and money. He considers these people to be blind to the objects of actual worth in their lives. The Larges seem to have embraced the tenets of fortune that Boethius illustrates.

The Larges, in the way they put their winnings to use, exemplify Boethius' writings on fortune. The couple has experienced both the results of fortune and misfortune. Violet is currently battling cancer. Many would become bitter upon experiencing that sort of calamity. However, the Larges do not curse their poor luck, as the ignorant men Boethius admonishes. Nor do they become greedy and horde their money for their own purposes. The Larges looked past the false value many in today's society place in large sums of wealth. They instead realized the value of each other's companionship, and the contentment of the simple life they led. They resisted the temptation of pursuing the transitory happiness that many think wealth offers. The Larges felt they had found happiness, and their lottery winnings would not offer them any further benefits. As Boethius commented in his work, "So why, mortal men, do you pursue happiness outside yourselves, when it lies within?" (27). The Larges seem to have found this inner peace that Boethius describes, and have placed their value in the aspects of their lives that Boethius would find satisfactory. Just as Boethius concludes that men should treasure both self and family, the only winnings that the Larges personally spent went to aiding their family. This generosity and contentment also relates to Boethius' thoughts on the concept of riches and the accumulation of wealth.

Boethius discusses the actual value and nature of riches and wealth at length in *The Consolation of Philosophy*. Boethius feels that money and riches as the ignorant majority of society deals with actually carries very little value. The possession of money provides very little value. Boethius actually states that a person's wealth does not actually belong to them. He instead holds that the actual value of wealth belongs to the wealth itself, rather than the owner. While this seems to be a very abstract idea, he explains it in a few simple analogies. In one of these analogies Boethius questions to whom the beauty of a jewel belongs. People buy and wear jewelry to draw attention to themselves, and to accentuate their own beauty. However, Boethius argues, "Their brightness of the jewel is the property of the jewels, not of the men who own them" (29). The stone's brilliance draws fascination for the stone, but really does not

describe the owner whatsoever, other than possible ignorance on the owner's behalf, Boethius might argue. In fact, Boethius contends that not only does the possession of riches not hold any value, but that it can even impose negative effects upon the owner. It does not necessarily cure their owner of wants or worries, but instead can serve as the catalyst for annoyance. When asked by Philosophy about the calming effects of wealth, Boethius answered, "I cannot recall ever feeling free from one worry or another" (45). The protection of immense riches creates a burden upon the wealthy man that the poor man never needs to feel. As the popular 1990's rapper Christopher "Notorious B.I.G." Wallace once said in his former hit single, "More money, more problems." Wallace was trying to convey the sense that life became much more confusing when he became rich, as compared to his earlier days of modest income. The concepts that Boethius uses to form his opinion on wealth are readily visible in the Larges' decision, especially his commentary upon generosity.

Allen and Violet Large clearly saw little value in the stockpiling of wealth, just like Boethius. The Larges saw no major personal uses to put the money towards. Many in the same position would elect to live out the remainder of the years in opulence, with nice clothes, diamonds, and the like. However, as Boethius would advise, they chose the life of simplicity they already possessed. They were not in great want, having saved up from years of labor. And, according to Boethius, wealth does not quell want anyway. With nothing they wished to spend the money on, and not perceiving any value in the simple possession of wealth, the couple gave the money to organizations that would work towards eliminating the needs that the accumulation of wealth would not satisfy. In doing so, the Larges supported the concept Boethius outlines that states that the amassing of immense wealth among the few isn't healthy, not only for the sole possessor of the wealth, but for the society as a whole. As Boethius states, "How restrictive and poverty-stricken are these riches, which cannot be possessed in their entirety by the many, and which do not pass to any single person without leaving the rest in want!" (29). When only a few collect the wealth of a society, and it is not put to any productive use, the whole society has suffered significantly. Instead, the Larges contributed the money to the collective good of society, thus bettering society and themselves. This act of generosity is exceptionally significant when viewed in the context of Boethius' commentary on the only true value that money

holds. Boethius writes, when considering generosity and the passing on of wealth, "Money only becomes valuable when bestowed on others by the practice of giving, thus ceasing to be possessed" (29). Boethius feels that the Larges' use of their riches is the only proper way to truly utilize the value of wealth. And it is for this reason that the Larges have come into the impressive amount of fame that they have experienced.

For an elderly couple in rural Canada who mostly keep to themselves, the Larges have come into quite a position of fame in the short time since their inspiring use of the lottery funds. The *Associated Press* article written about their decision has been carried in newspapers across North America in the past two weeks. As Allen Large told the new organization, "We're not used to all this attention ... We're just plain, old country folks" (CBS/AP). In a day when so many news stories are negative, and fame frequently assumes a sense of infamy, the Larges incredible generosity has inspired people. Boethius would feel that this fame was only natural considering their actions. Boethius comments at great length about fame in *The Consolation of Philosophy*, including a commentary on fame in the context of wealth. He writes, "Their sheen is more attractive when they are doled out rather than gathered in, for avarice always breeds hatred, whereas generosity brings men fair fame" (28). Boethius feels that people who generous with the wealth they have will gain not only a sense of righteousness, but fame as well. The Larges are a perfect example of this principle at work. They have become famous due to their generosity, and the fame has come with the utmost respect for their virtue.

The article about Allen and Violet Large is a testament to the fact that the principles Boethius set down to paper in *The Consolation of Philosophy* still endure today. Wealth remains just as much a false happiness and source of value today as it did in Boethius' time. Just as he wrote, "Wealth cannot ensure sufficiency, nor kingship power, nor high offices due respect, nor glory fair fame, nor pleasures joy" (53). The Larges actions proved that great wealth cannot cure want, and cannot provide happiness for the virtuous. Good fortune is not as beneficial as it seems, and most people overlook their true blessings. Allen and Violet, however, realize their blessings in each other and family, rather than riches and worldly possessions. Fame remains a fleeting goal that many men strive for, but only the generous and honest truly attain in the fairest sense. The Larges, on the whole, are a true

representation of the principles of Boethius' philosophy incarnate in the modern world.

Works Cited

Boethius, . *The Consolation of Philosophy*. New York: The Oxford Press, 1999. Print.

CBS/AP, . "COUPLE WINS \$11M; GIVES ALMOST ALL OF IT AWAY."

CBSNEWS.COM (2010): WEB. 13 JUL 2011. <[HTTP://WWW.CBSNEWS.COM/STORIES/2010/11/05/EARLYSHOW](http://WWW.CBSNEWS.COM/STORIES/2010/11/05/EARLYSHOW)

Struggling For An Answer to Capitalism: Evelyn Waugh and George Orwell's Pessimistic Approach

Joseph Wolf

The global economic crisis in the 1930's farther complicated the already unstable social and political upheaval of the era. The continued divergence of wealth and poverty created a dismal view of the future of the social world, while conflict loomed in the political. Themes in Evelyn Waugh's *Vile Bodies* and George Orwell's *Keep the Aspidistra Flying* appear to reflect a concern for capitalism as the basis for the English and global economic system. Both of the authors use satire and direct critiques of capitalistic qualities to illustrate this distrust, but Waugh and Orwell combat the issue from different ends of the socio-economic ladder. *Vile Bodies* satirically displays the impulsivity and irresponsibility of privileged English youths in their lack of concern for money, safety, or the wellbeing of others. Their actions show the impermanence of their class and economic situation in an unstable society and how these actions may contribute to their own downfall. Conversely, Orwell depicts lower classes in Gordon Comstock's fruitless battle against the "Money God", which leads him to a hellish life of poverty and sorrow. However, despite their critical stances, both authors fail to present a tangible alternative to the problems they explore. Instead, Waugh and Orwell present a prophesy of war and destruction as the inescapable, natural path for Capitalistic societies.

Evelyn Waugh was born into the upper-middle class, yet his involvement with or relation to individuals similar to The Bright Young Things in *Vile Bodies* is unclear. However, observation of such people led to his grim understanding of how their culture and lifestyle affected English society. Waugh's characterization of the careless youths in *Vile Bodies* draws from his impressions and exposure to the early 20th century Futurist art movement. In the 1909 *Futurist Manifesto*, F.T. Marinetti outlines the goals of the Futurists and the reasoning behind their art and actions. Coming into the new century, the futurists rejected the foundations of former art and society and embraced elements of speed and mechanization. Waugh's interest in these ideals is apparent in direct reference to them in *Vile Bodies*. The Manifesto states, "4. We declare that the splendor of the world had been enriched by a new beauty: the beauty of

speed... a racing automobile...is more beautiful than the Victory of Samothrace" (Marinetti). Although Waugh may not have completely agreed with all of the doctrines, the representation of these specific mechanical elements and rejection or ignorance of the past in favor of the present in *Vile Bodies* is apparent. However, to what ends does Waugh utilize these elements? It is entirely possible that he is mocking the Futurists when their glorious automobile's flawed nature ultimately causes Miss Runcible's death. Her death could also be in indication that the Futurist thinking is flawed and will lead to destruction if taken to an extreme. According to Brooke Allen, Waugh's writing style is similar to the Futurist style of art, "With a minimum of description Waugh succeeds in reproducing the aura of the recently modernized, mechanized city almost solely through his use of accelerated dialogue and truncated conversations" (321). Allen suggests that Waugh's writing is stylistically representative of the art form, which supports the idea that *Vile Bodies* could be an attempt to praise Futurist ideals. Based on the Manifesto, it is difficult to say exactly where the Futurists stand concerning capitalism. On one hand they praise, "the nocturnal vibration of the arsenals and the workshops beneath their violent electric moons," suggesting support of the power of the machine over man. While on the other, "want to exalt movements of aggression," such as organized labor. This complicates labeling Futurists as pro or anti capitalists definitively.

Another element linking Waugh to Futurism is the lack of political alignment that is characteristic in both Waugh's writing and modern artistic movements (Allen 319). This lack of ideological commitment separates Waugh from partisan conflicts and allows for fluidity in his art and personal life. He is shielded from an alignment that might cause rigidity in thinking and restriction of social interaction, which enables Waugh to have a unique perspective. Despite this, his conversion to Catholicism does suggest an inner connection with morality and human decency, which affect his perspective and outlook. Although Waugh had not yet converted at the time of this particular work, after his conversion he adhered to Catholicism as a "rock in the midst of social and moral decay",

which may explain some of his moralistic tendencies (Wiley 263). By observing the world around him, it is likely that the effects of capitalistic culture conflicted with what Waugh believed to be the morality of mankind even before this conversion. If this is the case, the pessimism in *Vile Bodies*, the depiction of the Bright Young Things, and the resolution in destruction could be Waugh's way of leveling things out moralistically. This pessimism is also reflected in the Futurist Manifesto. Section nine of the Manifesto states, "We want to glorify war – the only cure for the world" (Marinetti). Waugh seizes war as a way to extinguish indifference and fickleness of English capitalism instead of offering a solution in *Vile Bodies*. This seems to be an escape rather than a solution, and can be seen as one of Waugh's disadvantages.

The emptiness of the Bright Young Things is a critique of the capitalist society that supports them. Like a Futurist piece they are always on the move; searching for parties, social interaction, and new innovations, rather than politics and current events. Their irresponsibility and incompetency is shown in their lack of concern. Money, safety, and personal well-being are all thrown away in order to embrace this lifestyle of self-indulgence. They embody the constant movement and thus the beauty of continuous motion. Without such a system as capitalism, the Bright Young Things would neither have time nor the ability to enjoy the luxuries their life has to offer. As a direct result of their place in society they are free to live with no recourse for their actions. From their perspective the past is irrelevant and the future can be faced without fear. It seems that all the parties and extravagance that Waugh depicts show the wantonness of their lives, and comes to an apex when Adam, exasperated, says to Nina, "Oh Nina, *what a lot of parties!*" (Waugh 170). This irritation suggests that although the lifestyle may be fun, it does not achieve anything, and disrupts productivity and development. Adam seems to be one of the only people in the novel to come to this realization. This could be because his situation makes him a participant of the lifestyle but not a cause. His exasperated nature could be a voice of reason calling out for an answer to the point of it all.

Keep the Aspidistra Flying is George Orwell's satiric response to the same capitalistic society that Waugh criticizes in *Vile Bodies*. In his book, Orwell's aggression towards a capitalist economy, or at least consumerism, is presented through Gordon Comstock. On a personal vendetta against

the "Money God", Comstock quits a well-paying job and turns to a life of self-enforced destitution. This results in Gordon's endless pessimism, which is a device for relating Orwell's observations of a flawed and horrid world. Looking out the bookshop window Gordon sees poster advertisements:

Corner Table grins at you, seemingly optimistic, with a flash of false teeth. But what is behind the grin? Desolation, emptiness, prophecies of doom. For can you not see, if you know how to look, that behind that slick self-satisfaction, that tittering fat-bellied triviality, there is nothing but a frightful emptiness, a secret despair? The great death-wish of the modern world. (16)

This London street is brought to life devoid of redemption or hope; only one example of Gordon's dismal portrayal of his world.

Orwell, like Waugh, did not limit himself by subscribing to one ideology, however he did indicate that his novels are, "directly or indirectly against totalitarianism" (Lutman 149). Despite this overreaching theme in his novels this lack of ideological definition allows the freedom of open criticism for nearly anything, including Capitalism. Orwell makes it clear that Gordon in *Keep the Aspidistra Flying* is also ideologically ungrounded. He clearly does not support capitalism and does not appear to subscribe to much religious, political, or ideological doctrine. Gordon is described as having dabbled in Socialism at a young age, but he quickly outgrew it, "Every intelligent boy of sixteen is a Socialist. At that age one does not see the hook sticking out of the rather stodgy bait" (43). This direct rejection of Socialism is interesting because it hinders a clear alternative and possible solution to capitalism. This lack of identity leaves Gordon bitter and aggressive externally and internally.

Gordon's outward hostility is a manifestation of the inner battle that he subjects himself to. He left advertising and consumer society to find time and inspiration he felt resulted in true art, however Gordon finds himself in a more Hellish life than ever before. He needs money to create *London Pleasures* because of the comforts necessary to engage his work; Cigarettes, tea, and inspiration, which come from a well filled bank account and stomach. Instead Gordon only has one suit, a drab room with dregs in a contraband teakettle, and an unfinished manuscript. Realizing his inability to create whether well-off or destitute casts Gordon into downward spiral of despair. This indicates that either complacency or rejection of capitalism both end in misery, which culminates in inevitable death

and destruction that Gordon prophesizes. Whether this destruction is physical or moral is unclear, but analyzing George Orwell's stance on ideological and political matters, it can be inferred whether he truly believes that capitalism will result in the devastation Comstock thinks.

Although George Orwell may not have identified with any one political ideology it is still possible to trace the basis for his critiques of modern society. Gordon Beadle adheres to the fact that "the precise nature of his [Orwell's] political posture simply cannot be defined and analyzed within the context of any identifiably modern political ideology, party, or movement," but also suggests that Orwell's highly critical and political work and was the result of influence from the moral Victorians (278). This separated Orwell and enabled to freely and objectively critique the nature of his own time. The most dominant influence of the Victorian era seems to have been Charles Dickens, "in the formation of Orwell's social and political consciousness...Orwell shared Dickens's preference for a moral rather than ideological approach to social and economic injustice" (Beadle 289). Waugh and Orwell are very similar in their moralistic perspectives. Using morality and humanity as a compass to guide criticisms of the modern world, they both come to the same conclusion that capitalism has many morally objectionable traits. It is unclear whether it is the moralistic argument against capitalism that drives the hatred in Gordon Comstock or whether it is a general disgust for the system as a whole. It is likely that Gordon's contempt stems from the disparity between the indulgence of the wealthy and helplessness of the poor. However this moralistic element is somewhat confounded by Gordon's sexual encounters and self-indulgence when he comes into money. These actions are perhaps devices used by Orwell to indicate that morality and moral thinking are not a solution to the problems of capitalism and can offer no alternative. Beadle also notes that attachment to Victorian values could be the reason for Orwell's pessimism:

Orwell's novels, literary criticism, and social commentary may be read as a kind of Victorian critique of modern society. He departs from the Victorian intellectual tradition only in his pessimism, which was itself largely a product of the rapid and seemingly inevitable erosion of Victorian values and ideals in the twentieth century. (289)

This insight into where Orwell's sympathies and expectations for society lay is important because it may indicate where to look for good in his works. Gordon Comstock's teakettle is a possible symbol for the good in him because although not specifically Victorian, it exemplifies English traditions and ideologies that Orwell sympathizes with.

The aversion to capitalism in *Keep the Aspidistra Flying* brings up the same question that it does in *Vile Bodies*, what solutions are offered to the problems raised? The answer it seems is nearly nothing. Since Orwell pulls Gordon out of his long fight, nothing is really resolved. This uncertainty in the solution is reflected by Orwell's independent political stance. Beadle quotes Orwell as saying, "capitalism leads to dole queues, the scramble for markets, and war...collectivism leads to concentration camps, leader worship, and war. There is no way out of this" (291). Orwell focused his attention on criticizing the, "British class system, economic inequality, imperialism, and other aspects of the capitalist system" as other Marxists did, but he believed that "Marxism offered a false and dangerous solution to the evils of capitalism" (Beadle 294). This political indecision leaves both reader and critic little to go on when trying to find Orwell's overall answers to his own critiques. In the case of *Keep the Aspidistra Flying* it does not appear that there is a solution to this extremely difficult and complex situation. When faced with the new pressures of his relationship with Rosemary and a child on the way, Gordon reluctantly ends his war against the consumerist culture and resumes his previous job. Immediately after doing so, he feels an enormous weight lifted from his shoulders and becomes a bit less pessimistic, "once again, things were happening in the Comstock family" (248). At first glance this may seem like Orwell is indicating that capitalism is not necessarily evil after all. However, the intense portrayal of Gordon's entire fight suggests that Orwell does not want to accept this, but he has no alternative. It seems as though Gordon's return to his previous lifestyle indicates that capitalism has its flaws, but cannot be escaped or overturned by individual disruption.

When looked at side by side, the similarities between the messages of the two novels are clear. These two authors have created unique perspectives on capitalism by pointing out its flaws from both the top and bottom of the English hierarchy. Yet, both authors still appear to come to the same conclusion, despite their discontent with the

system. This inability to provide a solution to the problem could reflect a general feeling of helplessness in the English population as the global economy continued to plummet and the likelihood of war increased with the rise of fascism. The instability of global politics of the time is reflected by the pessimism of both Evelyn Waugh and George Orwell regarding capitalism, and uncertainty of other existing systems. With nothing else to effectively combat the injustices they saw with the capitalist society, they had to rely on their critiques to show the flaws in societal structures and hope that something better could come from them.

Works Cited

- Allen, Brooks. "Vile Bodies: A Futurist Fantasy." *Twentieth Century Literature*. Vol. 40. No. 3 (Autumn 1994): 318-328. Jstor. <<http://www.jstor.org/stable/441558>>. 12 December 2010.
- Beadle, Gordon. "George Orwell and the Victorian Radical Tradition." *Albion: A Quarterly Journal Concerned with British Studies* Vol. 7, No. 4 (Winter 1975): 287-299. Jstor. <<http://www.jstor.org/stable/4048298>>. 13 December 2010.
- Lutman, Stephen. "Orwell's Patriotism." *Journal of Contemporary History* Vol. 2, No. 2 (April 1967): 149-158. <<http://www.jstor.org/stable/259956>>. 13 December 2010.
- Marinetti, F.T. "The Futurist Manifesto." <mycourses.geneseo.edu>. 10 December 2010.
- Shreve, Emily. "From Vile Bodies to Bright Young Things: Waugh and Adaptation." *Evelyn Waugh Newsletter and Studies* 39.1 (2008): MLA International Bibliography. EBSCO. Web. 13 Dec. 2010.
- Wiley, Paul L. "Review: Evelyn Waugh." *Contemporary Literature* Vol. 9, No. 2 (Spring 1968) 261-264. Jstor. <<http://www.jstor.org/stable/1207498>>. 13 December 2010.

Kang Keqing: From Victor to Victim

Rebecca Pullano

Since the earliest recorded times in Chinese history, women had been considered inferior to men. However, the overthrow of the Qing Dynasty in 1911¹ marked the commencement of the destruction of traditional Chinese culture; this cultural change coincident with the growing seeds of Communism in the country, led to the incorporation of women into society in a seemingly more egalitarian manner. Yet, despite an outward appearance of equality, it is likely that providing women with greater rights was, in fact, a purposeful strategy of Mao Zedong and the Chinese Communist Party intended to gain support for Communism. In fact, one could argue that, although women had significantly more freedom after the fall of the Qing Dynasty than they had experienced in traditional China, the nation was still based in patriarchy, men and women were not treated equally. The façade of power that women assumed during the Communist Revolution is essentially a microcosm that reflects the larger picture of Mao Zedong's rule. Mao frequently used false appearances in order to secure his personal power as the head of China, especially throughout the Cultural Revolution. For example, during the Great Proletarian Cultural Revolution, which began in the mid-1960s,² many Communist Party members who were essentially responsible for helping Mao rise to power were removed from power on unreasonable grounds and shunned by Mao, who "preferred to use the masses more directly"³ to accomplish his goals because they were more willing to accept his beliefs without questioning him as higher party cadres were more likely to do.⁴ One important woman in Chinese history who exemplifies the notion of appearance versus reality under Mao's rule during the Chinese

Communist Movement that began shortly after the fall of the Qing Dynasty in the 1920s,⁵ and, later, during the Chinese Cultural Revolution, which lasted from 1966 to 1976,⁶ is Kang Keqing, known as the "Girl Commander" in the Red Army.

Traditional China was dominated by Confucian teachings, which stated that "women in the human order should be lowly and inferior like the earth,"⁷ but the dawn of modern China brought about reforms that were promising in terms of rights gained by women. After the collapse of the Qing Dynasty, the two factions of the United Front, the Kuomintang Party (KMT) and the Chinese Communist Party (CCP), found themselves engaged in a power struggle for leadership of China. The KMT was "the party of wealthy gentry and landlords, [and it] had no place for peasants or women in either their political or military organisations."⁸ The CCP, on the other hand, took advantage of the fact that women had been oppressed for so long. Rather than exclude women from joining the movement, "the CCP slowly rebuilt its popular support during the 1930s under [Mao Zedong's] rule, [and] its focus shifted from urban to rural recruitment, including the women of China's huge peasant class."⁹ During the revolutionary struggles, women played a vital role in the Communist cause, as they found themselves involved in production of resources and propaganda campaigns. In addition, according to historian Bernard Cook, "Communist women's groups focused on maintaining morale and enhancing women's status within the family and the Communist movement."¹⁰ Because Mao and the

⁵ Marc J. Belcher, *China Against the Tides: Restructuring through Revolution Radicalism, and Reform, Third Edition* (New York: The Continuum International Publishing Group Inc., 2010), 205.

⁶ William J. Duiker, *Contemporary World History, Fifth Edition* (Boston: Wadsworth Cengage Learning, 2010), 269.

⁷ Arvind Sharma, ed., *Women in World Religions* (Albany: State University of New York Press, 1987), 140.

⁸ Lily Xiao Hong Lee and Sue Wiles, *Women of the Long March.* (Sydney: Allen & Unwin, 1999), 79.

⁹ Bernard A. Cook, *Women and War: a Historical Encyclopedia from Antiquity to the Present, Volume 1.* (Santa Barbara: ABC-CLIO, 2006), 108.

¹⁰ Ibid., 109.

¹ Maria R. Haberfield and Ibrahim Cerrah, *Comparative Policing: The Struggle for Democratization* (California: Sage Publications, 2008), 16.

² William J. Duiker, *Contemporary World History, Fifth Edition* (Boston: Wadsworth Cengage Learning, 2010), 269.

³ R. Keith Schoppa, *Revolution and Its Past: Identities and Change in Modern Chinese History, Third Edition* (New York: Prentice Hall, 2011), 348.

⁴ Ibid., 346- 349.

CCP offered women greater rights, women were naturally drawn to support them, and holding leadership positions in the CCP made women feel as though they played an important role in the cause for which they were fighting. Kang, for example, chose to support the CCP because she was able to get more involved and was offered more personal and professional freedom. It is possible that, because the KMT was more exclusive in its membership, the CCP was able to gain momentum and emerge victorious from the struggle for leadership in China in the 1940s.

Later, after the CCP celebrated success and Mao emerged as China's leader, the women's movement continued as reforms were being made in the country. For example, a 1951 speech made by Teng Ying-chao, Vice-President of the All China Democratic Women's Federation proclaimed that the People's Republic of China (PRC), led by the CCP, made significant reforms that gave women "equal rights with men in political, economic, cultural, educational, and social life."¹¹ However, as evident from the experiences of female Red Army general Kang Keqing, Mao and the Communist party leaders did not intend to equate women's status to that of their male counterparts.

Kang Keqing was born in 1911 to a poor peasant family in Jiangxi Province, where she was adopted into a nearby farmer's family because her biological parents were too poor to raise Kang themselves.¹² Unfortunately, the family into which Kang Keqing was adopted became financially insecure, so when Communist organizers came to her village in 1926, her adopted family was enticed by the benefits Communism seemed to offer. In fact, her foster father was so enchanted with the ideas of Communism that he became the head of the village's Peasant Union.¹³ Kang was also captivated by these ideas and was repulsed by the "rapacious landlords and gentry of the KMT... [whose] soldiers and landlord militiamen committed atrocious acts of intimidation."¹⁴ In

¹¹ Teng Ying-chao, *The Women's Movement in New China* (All China Democratic Women's Federation, 1952), 1.

¹² Anne Commire and Deborah Klezmer, ed., *Women in World History. A Biographical Encyclopedia. Volume 8: Jab - Kyt* (Waterford, CT: Yorkein Publications, 2000), 440.

¹³ Lily Xiao Hong Lee and Sue Wiles, *Women of the Long March.* (Sydney: Allen & Unwin, 1999), 76.

¹⁴ Ibid., 77.

1926, Kang "joined the Communist Youth and was made captain of the Young Pioneers."¹⁵ In an interview with Helen Foster Snow, Kang referenced a massacre in February 1927 during which the KMT mercilessly murdered a thousand peasants, many of whom were not even Communists, in Kang's district. She said "'The killing of my comrades made me furiously angry and more revolutionary than before.'"¹⁶ Motivated by the Communists' cause, Kang joined the Fourth Front Army in 1928 to avoid an arranged marriage set up by her foster father. While working in Jinggangshan, where the Red Army base was located, Kang Keqing met Zhu De,¹⁷ the "father" of the Chinese Red Army.¹⁸ Although he was married when they met, his wife at the time, Wu Yulan, was captured and decapitated by Nationalist soldiers in 1929, and a few months later, Zhu and Kang were married. Kang admits that she was not in love with Zhu when she first married him, but she was drawn to his devotion to the Communists' cause and had great respect for him.¹⁹ As will be seen later on, this marriage allowed Kang Keqing significant benefits as a woman. She used her *guanxi*²⁰ as Zhi De's wife to her advantage, gaining opportunities that she would not normally have had as a woman.²¹

In 1930, the Fourth Front Army, pressured by the Nationalists, evacuated Jinggangshan and reestablished their base in Jiangxi.²² Kang travelled

¹⁵ Anne Commire and Deborah Klezmer, ed., *Women in World History. A Biographical Encyclopedia. Volume 8: Jab - Kyt* (Waterford, CT: Yorkein Publications, 2000), 440.

¹⁶ Nym Wales, *The Chinese Communists: Sketches and Autobiographies of the Old Guard: Book 1: Red Dust, Book 2: Autobiographical Sketches, Books 1-2,* (Greenwood Pub. Co., 1952), 213.

¹⁷ Lily Xiao Hong Lee and Sue Wiles. *Women of the Long March.* (Sydney: Allen & Unwin, 1999), 80.

¹⁸ Anne Commire and Deborah Klezmer, ed., *Women in World History. A Biographical Encyclopedia. Volume 8: Jab - Kyt* (Waterford, CT: Yorkein Publications, 2000), 440.

¹⁹ Dean King, *Unbound* (New York: Little, Brown and Company, 2010), 37.

²⁰ Guanxi describes the interplay of a complex network of personal and social relationships. It can be understood in terms of it not being just what you know, but also whom you know" [definition from "Glossary of International Trade Terms," *Credit Today*, <http://www.credittoday.net/public/2001.cfm>].

²¹ Lily Xiao Hong Lee and Sue Wiles, *Women of the Long March.* (Sydney: Allen & Unwin, 1999), 84.

²² Anne Commire and Deborah Klezmer, ed., *Women in*

with her husband and the Red Army and was engaged mainly in propaganda-related work, promoting the CCP and Red Army and recruiting people to join in the efforts.²³ In 1932, Kang and Wu Zhonglian were appointed to command the Women Volunteers of the Red Army where Kang served as regimental commander and Wu as political instructor. Although it seemed a major step to incorporate women into the Red Army in such a manner, it has been argued that it was simply a weak attempt for Zhu De to fulfill the Central Committee's decree to incorporate partisans into the army.²⁴ Women were not expected to be readily incorporated into combat positions in the Red Army, and it was rare for them to train in the Red Military Academy. For example, Kang was only one of two females to study in the academy; the other girl was expelled. Because of the rarity of women's participation in Red Military Academy, Kang's marriage to such an important official is what actually allowed her such an opportunity.²⁵ Although Kang did attend the Red Military Academy and had her hopes set on being a soldier, she still was not allowed to fight in battle. In fact, historians have noted that, "one can only surmise that allowing her to enroll in the first place was simple appeasement."²⁶ Clearly, women were not actually viewed as equal to men within the party, although the CCP claimed they were. Rather, women were given the appearance of power and were selected for specific positions that seemed important in hopes of appeasing them and making them feel as though they played significant roles in the party. These roles, however, were actually prescribed by men, and women did not truly have the freedom to choose *how* they wanted to participate. It appears that, although the CCP wanted to make it look as though they were extending opportunities to women and destroying the patriarchal nature of Chinese society, it was simply a means to an end in order to garnish support; they really had no intention of readily incorporating women into an egalitarian environment. In this case, Kang was exceptional because, since she was Zhu De's wife, she was

given special opportunities such as enrolling in the military academy, but it became apparent that the connections her marriage afforded her could only bring her so far, as she was still restricted from being a combatant and holding the positions that she personally desired.

In 1934, the Red Army ran into some complications. During the Fifth Campaign of the Communists, Chiang Kai-shek organized a blockade of 500,000 troops and an offensive attack of 400,000 troops against the Red Army, which was much smaller in number. Because of these actions, "the Communists decided to give up the Soviet districts and to save the Red Army by breaking the blockade and marching it to the north. This Long March began on October 14, 1934."²⁷ The mobile army consisted of 86,000 male soldiers and only thirty females, one of whom was Kang Keqing. Among the mere thirty women soldiers who participated in the main Army on the Long March, most were the wives of soldiers and important party leaders, such as Mao Zedong's wife, He Zizhen. The fact that Kang was one of only thirty women on the Long March shows that she and the other women were offered a rare opportunity; not all women were afforded the equality to men supposedly existent in the CCP under Mao. In Kang Keqing's case, it appears that this opportunity was given to her not only because she was married to a powerful man in the Party, but also because she remained childless. Kang viewed having children as detrimental to revolutionary work and blamed women for finding themselves in such a situation.²⁸ This comment shows that women in the CCP were still somewhat restricted to traditional roles; because Kang viewed having children as detrimental, it seems that once women did have children, they were again confined to tending to their children rather than participating in party work. Kang's lack of offspring allowed her to move further along in her party work because she was not constrained by childcare. Another comment by Kang furthers this notion. In an interview, when asked about "the so-called "Women's Problem," she promptly put a stop to that line of questioning: 'I don't care much about the women's problem; I always work with men, not women.'"²⁹ By recognizing her uniqueness from other Chinese women, however, Kang

World History. A Biographical Encyclopedia. Volume 8: Jab - Kyt (Waterford, CT: Yorkin Publications, 2000), 440.

²³ Lily Xiao Hong Lee and Sue Wiles. *Women of the Long March*. (Sydney: Allen & Unwin, 1999), 84.

²⁴ Ibid., 86.

²⁵ Ibid.

²⁶ Ibid.

²⁷ Nym Wales, *Inside Red China* (New York: Doubleday, Doran & Company, Inc, 1939), 57-58.

²⁸ Ibid., 153.

²⁹ Ibid., 188.

acknowledges how she differed from other women and was also admitting that women were not equal to men in China. Rather, her case was an extremely exceptional one, and even she was not considered equal to her male comrades. As the book *Women of the Long March* notes of Kang's position in the party after the Long March, "Few Long March women were so lucky, and, while Kang Keqing spoke of what she had achieved being made possible solely through her personal efforts and the goodness of the party, it is difficult to believe she would have been unaware of the privileges being the wife of Zhu De brought her."³⁰ Clearly, her status as the wife of such an important man in the Communist Party, coupled with the circumstance that she remained childless throughout her entire life, speaks to the fact that her case was unique and that most common women, who likely had children, were not usually afforded such freedoms.

Although Kang Keqing was given opportunities that most Chinese women were not during her revolutionary work, she was still limited in pursuing her professional interests to the same extent as a male. In fact, even her husband prevented her from fully achieving her goal of becoming a commander in the army. After successfully completing the Long March and settling into Yan'an, Kang "plunged into studies, hoping to become a ranking military officer."³¹ In Yan'an, Kang first studied at the party school and later joined the Anti-Japanese Political and Military Academy.³² She claimed that Zhu De supported her studies and encouraged her desire to become a commander, but evidence seemed to prove otherwise. When asked why there were no women's regiments in the Red Army, Zhu De claimed that the Red Army could not take them in because of disciplinary problems and inadequacies in their physical condition that prevented them from performing at the level most men could.³³ From these statements, it seems evident that Zhu De really had no intention of incorporating women

into the Red Army, despite Kang's belief that he, too, wanted her to participate in the military.

Although Kang wanted to believe that her husband was on her side, in the 1940s, Kang came to the realization that as her husband and the other men of the party did not truly share her perspective. Because the KMT and the CCP joined together in a united front to try to fight off the Japanese, in 1940 Zhu De was sent on a trip to Luoyang, on which Kang accompanied him, to meet with KMT officials to organize their actions.³⁴ It was on this trip that she first "became aware that what she expected from her life as a Communist, and what the Communists, in the influential form of Zhu De, expected of her might not coincide."³⁵ During dinner on the evening of her arrival, Kang was startled when she was addressed in a toast as Mrs. Zhu and even more surprised when her husband encouraged her to go along with it because that was the way they did things in the KMT. Later on the trip, she was further offended when Commanding Officer Wei of the KMT offered her a smaller pistol that he thought was better suited for a woman, and rather than defend her, Zhu encouraged Kang to accept his offer. Because she was so uneasy with the way her husband expected her to act, she requested to be excused from events, but was denied by Zhu De, who told her it was her obligation to put the revolutionary work before her personal comforts.³⁶ This experience of Kang is probably the clearest example of the "appearance versus reality" notion of the CCP's treatment of women. When it was convenient, the CCP was more than willing to liberate women and raise them up, but when this threatened what the CCP wanted to accomplish, women were expected to return to previous roles, inferior to men.

While Kang retained her dreams of becoming a combatant, Kang never saw the day in which this would be the case. After the Japanese surrendered to China in 1945,³⁷ Kang held various non-military positions in the Communist administration. For example, in 1946, Kang became vice-chair of the

³⁰ Lily Xiao Hong Lee and Sue Wiles. *Women of the Long March*. (Sydney: Allen & Unwin, 1999), 152.

³¹ Harrison Salisbury. *The Long March: The Untold Story* (New York: Harper & Row, 1985), 83.

³² Anne Commire and Deborah Klezmer, ed., *Women in World History. A Biographical Encyclopedia. Volume 8: Jab - Kyt* (Waterford, CT: Yorkin Publications, 2000), 441.

³³ Lily Xiao Hong Lee and Sue Wiles, *Women of the Long March*. (Sydney: Allen & Unwin, 1999), 155-156.

³⁴ Anne Commire and Deborah Klezmer, ed., *Women in World History. A Biographical Encyclopedia. Volume 8: Jab - Kyt* (Waterford, CT: Yorkin Publications, 2000), 441.

³⁵ Lily Xiao Hong Lee and Sue Wiles, *Women of the Long March*. (Sydney: Allen & Unwin, 1999), 166.

³⁶ Ibid., 167-168.

³⁷ Marc J. Belcher, *China Against the Tides: Restructuring through Revolution Radicalism, and Reform, Third Edition* (New York: The Continuum International Publishing Group Inc., 2010), 205.

Committee for the Protection of Children in the Liberated Areas. Two years later, Kang was placed on the Preparatory Committee for the foundation of the Democratic Women's Federation, and in 1949 was placed on the Standing Committee of the Women's Federation when it was officially formed. During her time on this committee, various reforms were passed that gave women more rights such as the Marriage Law of 1950. In 1955, Kang gained the position of secretary of the Democratic Women's Federation, which she held until 1957 when she became a vice-chair of the Federation of Women.³⁸ Although these were certainly positions of importance, Kang had requested in 1947 at the first meeting of the Central Committee's Women's Council to be released from the Women's Council and placed in a line of work that dealt with the military rather than with women.³⁹ However, this request was denied by all the delegates of the meeting, including Zhu De, who "publicly disclaimed her in his address...without mentioning names...[and] criticized that attitude of female cadres who were not willing to become involved in women's work."⁴⁰ Ironically, where her husband's position had previously allowed her more opportunities, "it is certain that her husband's high profile and the spectre of nepotism worked against her... Negative *guanxi* ensured she would never become a commander in the army."⁴¹ It is quite apparent from this situation that, although women were making grounds in the struggle to change Chinese society, they were still viewed as inferior to men, contrary to the equality that the Communist Party claimed to offer. Evidently, it was only on men's terms that women were granted rights, and it was only when it did not cause men of power to be inconvenienced that they would offer women more rights.

When the People's Republic of China was established, changes in China took place as Mao sought to purge his administration of anyone who went against him. Kang herself participated in the events of the Cultural Revolution to destroy Mao's "enemies," but later fell victim to the revolution. In

1957, Kang "unwittingly collaborated in the campaign to destroy Ding Ling, Peng Zigang, Pu Xixiu, and Liu-Wang Liming."⁴² However, both Kang Keqing and Zhu De became targets of attacks by the Red Guards in the mid-1960s. Both were attacked through propaganda posters; on his, Zhu was accused of being a "black commander," while Kang was targeted as being a "capitalist roader." Additionally, Kang was subjected to a "struggle session" during which she was attacked by the Red Guards and other Chinese citizens who were called to participate. From this, she came away deeply disturbed and resorted to spending most of her time with Zhu De in their apartment.⁴³ In 1976, the same year that Zhou Enlai and Mao Zedong died, Zhu De passed away. As Deng Xiaoping grew to power in the nation, Kang also began to assume more power, and in 1977, she was elected to the Central Committee of the Chinese Communist Party and the chair of the Federation of Women. After a long life of political work, in 1986, Kang retired from politics, and on April 22, 1992, she passed away.⁴⁴

Essentially, Kang Keqing's life experiences, although exceptional for a woman, provide evidence that women in post-imperial China were not necessarily given more power for the sake of increasing their personal influence or advancement. The women's movement seemed a ploy by ruling males to delude women into accomplishing the political agendas of the party in power. For example, it seemed as though the most popular position held by women in the CCP was that of a propagandist, which would serve their purpose quite well. Because women of traditional China had been confined to the home, having women out in the open, spreading the message of the Communist Party would certainly help convince people that Mao Zedong and his followers truly sought to change China and would encourage them to join in and support him. Although it is true and should be recognized that women had been given more rights than they had previously experienced, Chinese society under Mao and the CCP remained largely patriarchal and was clearly ruled and dominated by men.

³⁸ Anne Commire and Deborah Klezmer, ed., *Women in World History. A Biographical Encyclopedia. Volume 8: Jab - Kyt* (Waterford, CT: Yorke Publications, 2000), 441-442.

³⁹ Lily Xiao Hong Lee and Sue Wiles, *Women of the Long March*. (Sydney: Allen & Unwin, 1999), 219-220.

⁴⁰ Ibid., 220.

⁴¹ Ibid., 221-220.

⁴² Ibid., 239.

⁴³ Lily Xiao Hong Lee and Sue Wiles, *Women of the Long March*. (Sydney: Allen & Unwin, 1999), 237.

⁴⁴ Anne Commire and Deborah Klezmer, ed., *Women in World History. A Biographical Encyclopedia. Volume 8: Jab - Kyt* (Waterford, CT: Yorke Publications, 2000), 443.

The sad irony of women's rights in China is that their apparent advancement took place on men's terms and not for the improvement of women's stature or welfare, but rather to accomplish the motives of Mao and the CCP. Ultimately, however, the treatment of women is just one example of how Mao utilized deception in many aspects of his long rule of China. In the mid-1950s, for example, Mao encouraged intellectuals to criticize the government's actions during the Hundred Flowers campaign, but he ultimately ended up viewing the criticism as a threat to his power and launched a campaign to eradicate any perceived opposition.⁴⁵ Similarly, in later years during the Cultural Revolution, Mao expelled the very men who helped him gain power because he felt threatened by these party cadres. With the support of the masses, who were blinded by Mao's promises for a better China, he was able to strip these men and women of their authority and leave them politically and socially immobilized.⁴⁶ Macrocosmically, it appears as though Mao's attempted advancement of China as a whole was actually intended to accomplish the motives of his personal agenda and solidify his personal power, and women in China at this time were just one group offered a façade of power for the sake of furthering Mao's goals.

Bibliography

Belcher, Marc J., *China Against the Tides: Restructuring through Revolution Radicalism, and Reform, Third Edition*. New York: The Continuum International Publishing Group Inc., 2010.

Commire ,Anne and Deborah Klezmer (editors), *Women in World History. A Biographical Encyclopedia. Volume 8: Jab – Kyt.* Waterford, CT: Yorkin Publications, 2000.

Cook, Bernard A., *Women and War: A Historical Encyclopedia from Antiquity to the Present,*

⁴⁵ "Hundred Flowers Campaign'." New World Encyclopedia.

⁴⁶ R. Keith Schoppa, *Revolution and Its Past: Identities and Change in Modern Chinese History, Third Edition* (New York: Prentice Hall, 2011), 351-359.

Volume 1. Santa Barbara: ABC-CLIO, 2006.

Duiker, William J. *Contemporary World History, Fifth Edition*. Boston: Wadsworth Cengage Learning, 2010.

"Glossary of International Trade Terms," *Credit Today*.

Haberfield, Maria R. and Ibrahim Cerrah, *Comparative Policing: The Struggle for Democratization*. California: Sage Publications, 2008.

King, Dean. *Unbound*, New York: Little, Brown and Company, 2010.

Lee, Lily Xiao Hong and Sue Wiles, *Women of the Long March*. Sydney: Allen & Unwin, 1999.

Salisbury, Harrison. *The Long March: The Untold Story*. New York: Harper & Row, 1985.

Schoppa, R. Keith, *Revolution and Its Past: Identities and Change in Modern Chinese History, Third Edition*. New York: Prentice Hall, 2011.

Sharma, Arvind ed., *Women in World Religions*. Albany: State University of New York Press, 1987.

Wales, Nym. *The Chinese Communists: Sketches and Autobiographies of the Old Guard: Book 1: Red Dust, Book 2: Autobiographical Sketches, Books 1-2*, Greenwood Pub. Co., 1952.

Wales, Nym, *Inside Red China*, New York: Doubleday, Doran & Company, Inc, 1939.

Ying-chao, Teng. *The Women's Movement in New China*, All China Democratic Women's Federation, 1952.

Young, Helen Praeger *Choosing Revolution: Chinese Women Soldiers on the Long March*, Chicago: University of Illinois Press, 2001.

Chinese Expansion into the Western Hemisphere

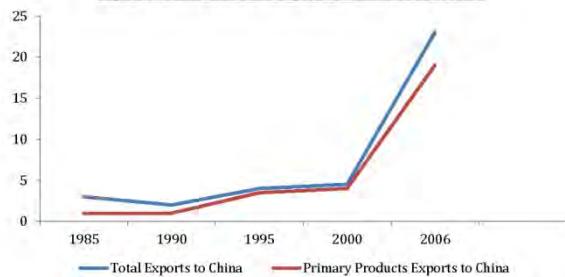
Steven Carlson

Introduction

Since the market-oriented reforms of Deng Xiaoping, China has consistently been increasing its economic interconnectedness with the outside world. The amount of goods flowing both to and from China is enormous and continues to grow. In 2010, China surpassed Japan as the second largest economy in the world. With an annual GDP growth rate averaging at approximately 10 percent, and a highly adaptable, growth-oriented Chinese Communist Party (CCP), one may only roughly estimate the limits of China's future economic strength. This growth has caused China to seek certain goods from abroad for new economic developments in both exportation and domestic consumption. In recent years, China's economic growth has led it to engage in meaningful contact with Latin America. As a net importer of primary products, China's chief reason for involvement in Latin America is for its primary products. China's expansion abroad is driven by economic reasons, with international politics being a means to greater economic gain. As with its dealings in other countries, China approaches Latin American countries with pragmatic, state-specific strategies, playing to the economic and political climates of its partners in order to access much needed natural resources. China's foreign policy is driven mainly by economic concerns and with little ideological content, making China a new and powerful player in the international scene.

Roughly around the turn of the century, exports from Latin America to China began to climb dramatically (see fig 1). This growth has continued into the present. The majority of these exports are primary products, which include minerals, fuel, and agricultural products.

Figure 1ⁱ: Latin American Exports to China, 2005 Dollars

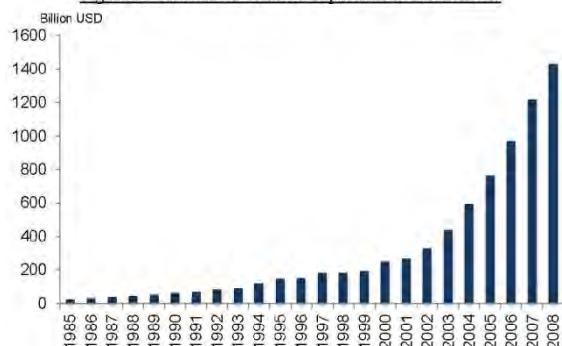


therefore a net importer of minerals, fuel, and agricultural goods. At the same time, China's economic growth has also been export-driven (see fig 2). China's growth in exports has exceeded that of every other nation on earth. With an export-driven economy, China has been able to make leaps and bounds in its economic development.

Economic Drivers of China's Natural Resource Demand

In the early stages of China's rapid growth, products that China exported were mostly cheap, low value-added goods. More recently, however, China's exports have become high-tech, high value-added products. In 2006, China surpassed the United States, the EU-27, and Japan to become the largest high- tech exporter in the world. Its high-tech exports grew 33 percent annually from 1995 to 2008, much faster than ⁱⁱⁱ the growth of overall exports. These new products necessitate vast quantities of primary products including minerals and fuel, the products that Latin America has in abundance, and the leading cause for China's aggressive economic expansion into the region.

Figure 2^{iv}: Growth of Chinese Exports in \$USD Billions



The Chinese leadership works to maintain its ambitious economic growth even with its relatively low per capita levels of natural resources. Over the past 18 years, China has become a net importer of almost all major natural resources. In 1993, China became a net importer of oil. From 1995 and 2005, its oil imports multiplied by a factor of five, becoming the second largest consumer of oil. During the

same time period, imports of iron ore grew 570 percent, copper by 738 percent, cobalt by 4,145 percent, and aluminum by 2,246 percent China has looked to the developing world for these products. China has developed into an exporter of high-tech products through its investment in education.^{vi}

In the decade ending in 2004, enrollment in higher education rose by a factor of 5, exceeding 13 percent of the college-age population.^{vii} Although percentagewise this is not a huge number, in absolute terms it means that China has over 2.4 million graduates per year, showing no signs of ebbing. Furthermore, China's education is dominated by students of science and engineering- two subjects crucial to the development of high-tech industries and key to China's future growth.^{viii}

Investments in infrastructure and research and development are also signs that China's growth will be long-lasting. China spends at least 1.2 percent of its GDP on research and development, more than any Latin American country.^{ix} By 2020, the CCP has a formal goal of making it 2.2 percent of GDP, with its main objective to have 60 percent of the country's economic growth driven by science and technology by 2020.^x China's investment in fixed assets is about 40 percent of GDP, twice that of the Latin American average.^{xi} This investment means better roads, better access to electricity, and better telephone land lines; all crucial to future economic growth.^{xii}

Its investment in research and development and internal infrastructure are powered by its access to cheap bank credit. China's people save extraordinary amounts of money. Their savings rate is approximately 50 percent of China's GDP. This offers the Chinese government access to cheap credit. Provided with this capital, China can develop forward looking economic developments.

China's demand for foreign natural resources will continue along with its economic advancement. China's booming railway construction, shipbuilding, and automobile manufacturing sectors are all natural resource-intensive industries, necessitating primary products such as metals, rubber, fuel, etc. Ownership of domestically produced automobiles has sky rocketed. In China, car ownership more than doubled between 2004 and 2009, from 12 percent of the population to 28. One report concludes that it is possible for car

ownership to top 50 percent of the population by 2015.^{xiii} Growth in Chinese domestic consumption continues to drive its need for foreign natural resources, making China and Latin America complementary partners.

With this economic growth has come a higher standard of living for most Chinese, who now demand more agricultural products from abroad. China is by no means a food insecure country; however, there are some agricultural products that China cannot sufficiently supply to its own people. While China is a major producer of many agricultural products, overall it is a net importer of bulk food commodities. China has become a major market for imports of oilseeds (ie. soybeans), vegetable oils, oil meal, and grain.^{xiv}

China's Access to Latin American Natural Resources

Latin America offers vast amounts of agricultural products, metals, and fuel. Latin America's agricultural land per capita is three times greater than China's while forest land and fresh water per capita is more than five times greater. Latin America produces about 50 percent of the world's soybeans, over 20 percent of global refined aluminum, 20 percent of the world's zinc, 30 percent of the world's oilseeds, 40 percent of the world's copper^{xv} and at least 14 percent of the world's petroleum.^{xvi}

Latin America's economic environment is complementary to China's. China has increased its exports of high-tech products from approximately 3 percent of total exports in 1990 to approximately 33 percent in 2004 (see fig 3). Meanwhile, Latin America's economy has continued to rely on the exports of primary products. Since China's new high-tech industries are natural resource-intensive, Latin America's economy is complementary to China's. China's imports from Latin America have shown tremendous growth (see fig 4).^{xvii}

Figure 3^{xxviii}: Technological Content of Exports, China and Latin America, 1990 and 2004

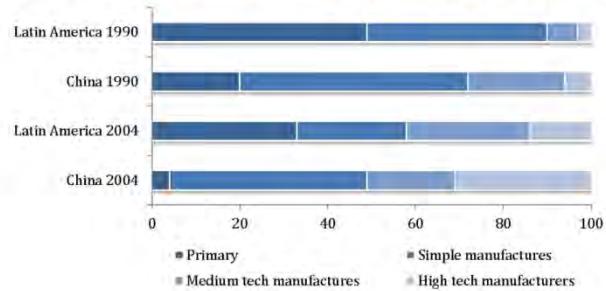
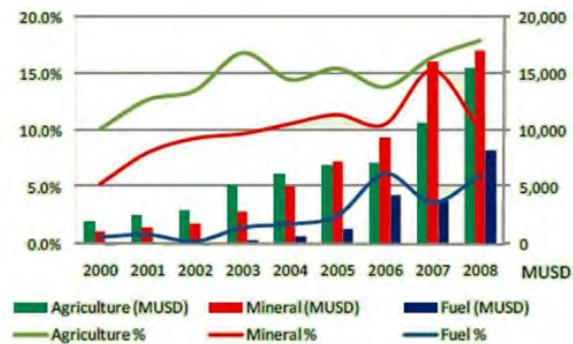


Figure 4^{xxix}: China's Imports from Latin America: Value & Percentage of World Total



These natural resources, however, are not found in significant quantities in every Latin American country. The bulk of the exports from Latin America to China has been heavily concentrated in the few countries that contain the primary products that China demands. In 2008, while the total volume of Latin American exports to China accounted for only 5.6 percent of the region's overall exports, for a number of countries in South America, China has become their top export customer, including Chile and Brazil's. China has become the second most important export customer for Argentina, Peru,

^{xx} Costa Rica, and Cuba. By 2008, just 10 commodities in only six Latin American countries accounted for 74 percent of the region's exports to China.^{xxi} These same 10 commodities made up 91 percent of the region's overall

commodity exports to China.^{xxii} China's expansion in trade with Latin America has primarily involved those countries with large amounts of natural resources.

These states that China is engaging are adapting their economies to meet China's demands. For example, China has had a profound impact on Latin American soybean production. Soybeans and soybean-derived products are an important part of the Chinese diet and are used for soy flour, soymilk, soybean oil, tofu, textured

^{xxiii} vegetable proteins, and soy sauce. China consumes about 40 million tons of soybeans annually, and it is predicted that China may soon import over 50 percent of the world's soybean production. Brazil and Argentina are two of the largest soya producers in the world. In 2009, China surpassed the U.S. as Brazil's most important trading partner. Argentina now devotes over half of its farmable land to the production of soybeans.^{xxiv} China's economic demands now determine economic production in large Latin American economies.

While the physical distance between Latin America and China is long and inhibitory to trade, developments are in place to facilitate commerce. Already under construction is a \$5.25 billion expansion to the Panama Canal that will add an extra lane and two sets of locks, tripling cargo capacity through the canal, and allowing China to have easier access to Latin America's eastern countries. China will be able to use its large, more efficient cargo vessels to transport raw materials from northeastern Latin America, particularly Brazil for iron ore and Venezuela for oil. Construction of the additional parts to the canal are projected to be completed by 2014.^{xxv} Another possibility for linking the eastern Latin American economies to the Pacific is through a proposed "dry canal" through Colombia. China and Colombia are in advanced talks to develop a \$ 7.6 billion 791 kilometer railway through Colombia.^{xxvi} These talks represent China's aggressive push for greater access to Latin American natural resources.

China's Engagement with Latin America

China engages natural resource-rich Latin American trading partners regardless of their political environment or the ideologies of their governments. Its interactions with Latin American countries are dependent on economic opportunities. In market-oriented countries, it has signed several Free Trade Agreements (FTAs). For example, China signed a FTA with Chile in 2006, with Peru in 2009, and with Costa Rica in 2010. China has also engaged Colombia for a FTA.^{xxvii} In socialist, revolutionary countries found in Latin America, China has engaged them in state-level economic developments. The leaders of Venezuela, Bolivia, and Nicaragua

have all increased their anti-American rhetoric, turning to China as alternative leader in the region. China has not shied away from business opportunities with the revolutionary governments of these countries. Its opportunistic play on Latin America's domestic politics is conducive to its economic advancement in the region.

Other Chinese Concerns in Latin America

Besides China's interest in Latin American exports, it also works to open markets for the sale of Chinese goods. Although Chinese firms seek to sell their products to any country they can, Latin America is particularly important for a number of reasons. First, China would like to diversify its export base. China relies mainly on developed nations for the purchasing of their goods. With the economic crisis of 2008 and the contraction of consumption by developed nations, China recognizes the need to diversify its export base to middle-income economies as a way to further guarantee the advancement of its economic

^{xxviii} goals. The growth of Chinese goods being imported to Latin America coincides with and is comparable to that of Latin American goods exported to China (see fig 5).

Latin Americans are important consumers for knock-off products of things like designer clothing, pirated music CDs, DVDs, footwear,

^{xxx} and other products. While this kind of trade is illegal, many customs authorities are not able to block the entry of these products or their sale, because these organizations are often under resourced, corrupt, or overwhelmed by the

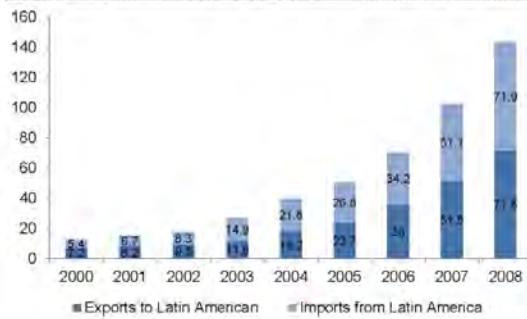
^{xxxi} volume of products. Products like computers, motorcycles, and cars from China are much cheaper than those from the developed world. While these products are often lower quality, many Latin Americans cannot afford products from developed countries. Therefore, "Latin America, with a population of over 500 million and an economy of nearly \$3 trillion is an

^{xxxii} attractive market for Chinese products." By selling low quality, and even illegal goods, the Chinese make economic gains from exporting to Latin America.

China engages countries for economic purposes, with ideological concerns only being a secondary matter. However, there is one exception to this rule. This exception pertains to the extension of its "One China Policy," or the

diplomatic isolation of Taiwan. Some countries in the region still recognize Taiwan's independence.

Figure 5^{xxxiii}: China's Bilateral Trade with Latin America (\$USD Billions)



However, these countries are not large natural resource producers, so they are not economically significant to China. The countries that recognize Taiwan's independence are mostly located in Central America and the Caribbean, and do not serve as highly important natural resource providers to the People's Republic of China (PRC). China uses its economic might to gain adherents to its policy. For example, it gave \$11 million in aid in 2005 to the Commonwealth of Dominica and a \$100 million grant over the next six years in return for its recognition of the PRC. The isolation of Taiwan in Latin America is China's main political concern; however, it does not significantly effect China's economic goals, and its economic expansion into Latin America is ^{xxxiv} not to isolate Taiwan. China's government's main goal in Latin America is for natural resources.

China's Government Facilitates Expansion

China's government works to facilitate the acquisition of foreign natural resources while maintaining oversight of this crucial sector of the economy. The CCP has reacted to China's dependence on foreign natural resources in two ways. Firstly, it has opened its markets to foreign suppliers by lifting quantitative restrictions on almost all commodities and by curtailing import

^{xxxv} subsidies. Secondly, because they realize the strategic importance of these commodities, leaders in the government have worked to maintain control over the production and trade of these natural resources through State Owned Enterprises (SOEs).

The Chinese government works actively to increase China's natural resource acquisition throughout the world. It supports and advances SOEs specializing in natural resources in order to

keep oversight and control over the trade of these commodities. The Ministry of Foreign Affairs and the Ministry of Commerce oversee the execution of national policy governing natural resources, with the exception being over oil. The Ministry of Foreign Affairs and the Ministry of Commerce are funded by the Ministry of Finance, which provides cheap credit for the companies' acquisition of natural resources. The oil trade is supervised by the Energy Leading Group. Because of the increased security importance of oil, the Energy Leading group is chaired directly by the Prime Minister. China's government has a coordinated strategy for natural resource acquisition.^{xxxvi}

The Chinese government knows that natural resources are connected to national security and international politics. China has been discriminated against in its quest for natural resources in Canada and the United States because of both security and nationalistic

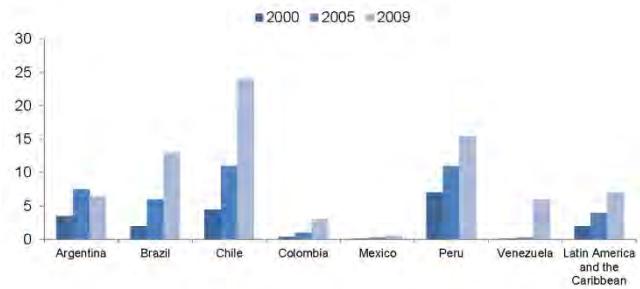
reasons.^{xxxvii} The natural resource markets are strategic and can be volatile in times of international crisis. It therefore behooves all governments to work to ensure a diversified portfolio of natural resource access. China continues to diversify its sources of natural resources.

China's foreign policy is sometimes controversial because of its seeming lack of moral guidance. In Africa, it is notorious for doing business with the worst of governments, in countries where most Western companies will not even enter because of humanitarian concerns. An example of this is in Sudan, where China has millions of dollars invested in the oil sector, and has worked to ensure "stability," by blocking U.N. military action to stop the genocide occurring there. While the West is steering clear of Africa because of humanitarian and political concerns, China has swiftly moved in for natural resources. It is opportunistic in Africa, working to win natural resources where the West does not have a strong presence. China uses this same pragmatism in Latin America, and is willing to adapt its strategy to each country's political and economic concerns. Driven by its desire for primary products, China engages each trading partner opportunistically. Where China can access natural resources through FTAs, China does so, but where state-level development projects are the only means, China engages on these terms. When it comes to natural resource security and

acquisition, China's does not take moral, political, or ideological concerns into account, making it a rational, powerful player in the international scene.

Fortunately, in recent years, Latin America has not been the scene of large-scale human rights violations and genocide, as has been the case in Africa. However, Latin America has a wide range of ideologies, both economic and political, in the governments of the particular nations. As a rational actor in the international scene, China engages these states with expedience, with the primary purpose being natural resource acquisition. While there are numerous examples that demonstrate China's pragmatic, state-specific strategies, China's dealings with Chile and Venezuela are two exemplary cases of its adaptability in the region. These two countries have seen the greatest percent increase in their exports to China (see fig 6). They have very different governments, economic systems, and reasons for increased trade with China. Nevertheless, China has made extraordinary gains in accessing their most important natural resources.

Figure 6^{xxxviii}: Latin American Exports to China as a Percent Share of Total Exports



China's engagement with Chile

The historical relationship between China and Chile and the recent uptake of Chilean natural resources demonstrate how China's foreign policy in Latin America is based mainly on its demand for primary products, the most important of which is copper. China has supported Chile politically in international organizations and its leaders have engaged in numerous meaningful state visits to Chile. Playing to Chile's tradition of free market economics, China engages Chile with free market initiatives. At the same time, state level action has been taken in many instances for the production and sale of Chilean minerals to China. China has successfully gained profound interconnectivity with the Chilean

economy, gaining access to Chilean minerals by political and economic advancement.

Chile's geography and abundant mineral wealth make it a logical trading partner for China. Lining the west coast of South America, Chile's ports serve as intermediate stops for many goods flowing between the continent's interior; including Argentina, Brazil, Bolivia, Paraguay, and Uruguay; and Pacific-bound destinations like China. Ninety-eight percent of Chilean exports come from nine primary-

^{xxxix} product sectors. Chile holds approximately 17 percent of the world's copper reserves and is the world's largest producer of the metal. In the first half of 2007, over 83 percent of ^{xli} Chile's exports to China were copper. Chile's geographic location and vast natural resources make it a crucial partner.

Chile was the first country in South America to establish diplomatic relations with China in January of 1971. Recognition of the PRC came during the leadership of the Marxist president Salvador Allende, who looked to China's Communist regime as a potential political and economic partner. The two nations engaged in trade; Chile exporting mining products, fishmeal, and agricultural products, and China selling cheap manufactured goods. While formal diplomatic relations between Chile and China began during the presidency of this leftist, China's reasons for its relations with Chile were less about their shared ideology than about China's ^{xlii} access to Chile's resources.

With the fall of the Marxist regime in Chile, China demonstrated its apathy for Chile's political ideology. In 1973, Allende was ousted and the anti-Communist military dictator Augusto Pinochet took power, establishing a radically anti-Communist government. During this phase in Chile, the two nations retained economic and diplomatic ^{xliii} relations with low ideological content. The new regime implemented free-market reforms in an attempt to curb inflation and spur economic growth, while at the same time stifling civil liberties and violating human rights. The target of this aggression was usually Marxists and others who had left-leaning tendencies. Pinochet left Chile with a legacy of anti-Communistic, free market oriented reforms.

Up until the turn of the century, China and Chile's relations maintained cordial but ^{xliii} lukewarm. Changes in Chile's internal government to a more centrist, liberal democracy had little bearing on the Sino-Chilean relationship. Besides Chile's support for China's membership in the World Trade Organization in 1999, the 1990's showed little change in the interaction between these two ^{xliv} countries.

Since the 1990's, however, China has engaged Chile bilaterally and supported Chile in international organizations, coinciding with a stark growth in China's imports of Chilean goods. This correlation is no accident, as China uses international politics for economic gain. Between 1997 and 2006 Sino-Chilean bilateral ^{xlv} trade increased by 641 percent. President Hu's visit to Chile in 2004 during the annual Asia Pacific Economic Cooperation (APEC) summit demonstrated the increasing importance of Chile and China's economic relationship. China supported Chile's bid for a nonpermanent seat on the United Nations Security Council in ^{xlvii} 2003-2004. By 2007, the PRC became Chile's most important export customer, and by 2009 the PRC surpassed the United States as Chile's most important trading partner ^{xlviii} overall. China's political support of Chile internationally coincides with increased trade between the two.

Chile's conservative, free trade tradition has served as an opportunity for China to increase its economic activity with them. Free trade in Chile is a result of the market reforms of the post-Allende era. The "Chicago Boys," a group of Chilean economists educated at the Chicago School of Economics, influenced Chile's economy by starting free-market oriented reforms that have made Chile a pioneer in the region. Since 1985, Chile has had an average GDP growth of approximately 6.9 ^{xlviii} percent. Many attribute Chile's economic success to their free trade. In light of Chile's free trade tradition, China has engaged Chile with free trade initiatives.

China increased its access to Chilean natural resources through the groundbreaking Sino-Chile FTA that went into effect in 2006, guaranteeing increased trade and investment between the two countries. In October of that

year, China lifted tariffs on 2,834 products imported from Chile, including its most important commodity, copper. Chile gave duty-free status to 5,891 products from China.^{xlix} Chile's exports to the PRC jumped 140 percent in 2006, the first year in which the FTA went into effect.^l The FTA represents China's first ever FTA with any nation outside of Asia. By 2009, trade between the two nations reached record levels at \$16.5 billion.^{li} In signing this FTA, the Chinese were able to guarantee greater access to the primary products that it needs from Chile.

China has also engaged Chile on a state level basis in order to better control the flow of Chilean natural resources. In February 2006, China and the Chilean copper company CODELCO signed an agreement where China agreed to purchase Chilean copper at a fixed price, providing \$550 million in advance to help the Chileans increase their output in order to keep up with growing Chinese demand.^{lii} This agreement has an estimated value of about \$2 billion.^{liii} Through state level deals, China is able to better coordinate from government to government the flow of natural resources.

The growth of China's influence on Chile's copper industry has caused concerns of Chinese dominance and possible dependency for some in Chile. A controversial provision of an accord between CODELCO and China Minmetals, the Chinese state-owned mineral company, was an option for them to buy up to 49 percent interest in a new Chilean mine, Gabriela Mistral. To many Chileans, metals and minerals are a national patrimony and the possibility of foreign dominance of them^{liv} touches a nationalistic sensitivity. When in 2007 China Minmetals expressed its interest in purchasing 49 percent of the mine, workers in one of CODELCO's leading mines protested,^{lv} causing a sharp reduction in its production. The Chilean government decided to forbid the Chinese from exercising its option. This growing concern by many in Chile shows the profundity of China's presence.

China's trillions of dollars in reserves allowed it to continue its natural resources buying spree from Chile even during the Financial Crisis of 2008. In 2008, Chinese

imports from Chile rose 204.3 percent.^{lvi} This rise was driven by copper. In that year, China's imports of refined copper and alloys increased by 640 percent, imports of copper waste and scrap went up 716 percent, imports of copper ores went up 732 percent, imports of copper wire went up 300 percent, and imports of other copper products rose at comparative levels. While the U.S.' and Europe's economies were contracting, China's ability to buy Chilean copper increased.^{lvii}

The relationship between these two countries has been so lucrative, that in 2010 the two states began a new phase in the FTA to expand their free trade to include greater access of service suppliers into a wide array of sectors including mining and geological prospecting. In August 2010, Mario Artaza, head of Asia and Oceania affairs at the Chilean foreign ministry said, "since 2005, China and Chile have walked on a joint path^{lviii} toward free trade." China's imports of Chilean primary products have increased dramatically since the turn of the century. Engaging Chile with bilateral FTAs and state-level deals has given China greater access to Chile's natural resources, especially copper. China has become an integral part of Chile's economy. China has achieved economic interconnectivity with Chile because of its adaptable strategies of bilateral engagement.

China's engagement with Venezuela

Another example of China's expedience in its natural resources acquisition in Latin America is in its relations with Venezuela. China's expansion in bilateral trade and investment into Venezuela corresponds with the leadership of the revolutionary, anti-American president Hugo Chavez. Chavez offers China access to Venezuela's petroleum as a way to increase economic and political relations between the two countries. This is driven by his desire to lower Venezuela's dependence on the United States for consumption of its oil. With a foreign policy in Latin America based primarily on its quest for natural resources, China takes advantage of the political situation of Venezuela in order to access Venezuela's vast reserves of petroleum.

Although Venezuela does have large

reserves of other natural resources such as aluminum, iron, gold, and coal, by far its most important and abundant natural resource is petroleum. Venezuela has the second largest reserves of oil in the world, just behind Saudi Arabia Brief.^{lx} Venezuela is the seventh largest exporter of oil in the world.^{lx} Oil makes up 80 percent of Venezuela's total export revenue and contributes to approximately 50 percent of the central government's income.^{lxii} It is responsible for about one-third of the country's GDP.^{lxiii} Including the deposits of hard-to-reach oil deposits in the Orinoco belt, Venezuela has approximately 211 billion barrels of proven oil reserves.^{lxiv} The majority of Venezuela's oil is consumed by the United States.

Venezuela's leadership before Chavez showed little interest in economic and diplomatic relations with China, and there was little bilateral activity. However, Venezuela granted China diplomatic relations in 1974. The first visit to China by a Venezuelan president did not occur until 1981.

Likewise, the first visit of a Chinese premier to Venezuela did not occur until 1985, 11 years after Venezuela established diplomatic relations with China.^{lxv} In 1996, Premier Li Peng visited Venezuela. With only three high-level visits up until 1999, relations between the two nations were of little consequence.^{lxvi}

The political relationship between China and Venezuela increased dramatically with the election of Hugo Chavez. Immediately after winning, Chavez made his first visit to the PRC in October 1999, and repeated visits six more times by 2011. Chavez' statist, anti-free-enterprise stance and desire to increase political and economic ties with China have strengthened the relationship of the two. In 2001, China named Venezuela a "strategic partner," signaling its desire for increased cooperation and bilateral progress.

Because of Chavez, Venezuela has become an undesirable market for many Western companies. Starkly anti-capitalist, in 2006 he nationalized all of the oil fields in the Orinoco region that were owned by foreign companies, resulting in an increased ownership stake from 40 percent to 60 percent

for the central government.^{lxvi} The American companies ConocoPhillips and ExxonMobil, Venezuela's two biggest partners in heavy oil development at the time, decided to pull out of the country all together and settle it by international arbitration.^{lxvii} Seeing opportunities left by these private companies, China has worked to take advantage of the situation.

Chavez has clearly positioned himself as both anti-American and pro-Chinese. In 2006, at a U.N. conference in New York, Chavez was quoted as saying "The American empire is doing all it can to consolidate its hegemonic system of domination, and we cannot allow him [President George W. Bush] to do that. We cannot allow world dictatorship

^{lxviii} to be consolidated."^{lxix} President Chavez blames the attempted pro-United States coup in 2002 on the United States government, and was angered that the United States diplomatically supported the usurping regime. Since the failed coup, Chavez' anti- American sentiment has grown. Chavez desires to decrease Venezuela's dependence on the U.S. as its most important importer of oil and would like to increase the prominence of the newly rising power, the PRC. In a 2004 state visit to Beijing, Chavez said "You should realize we've been producing and exporting petroleum for more than 100 years, but they were 100 years of domination by the United States. Now we are free and at the disposal of

the great Chinese nation."^{lxix}

Statements like these give China confidence for aggressive development in the petroleum sector with Venezuela.

However, there are significant obstacles to Chavez' desire to meaningfully decrease Venezuela's dependence on the United States. First is the U.S.' close proximity to Venezuela. Just across the Caribbean, sending petroleum to the United States makes the most economic sense. That contrasts to China's far distance and the fact that in order to reach China's shores, ships must either go through the Panama Canal, around South America, or around Africa. Today, about 60 percent of Venezuela's oil exports go to the United States, making Venezuelan oil imports 11 percent of all oil imports to the United

^{lxx}
States. Petroleos de Venezuela SA (PDVSA), Venezuela's state owned oil company owns five oil refineries in the United States and has partial ownership in four through its U.S. subsidiary, CITGO.^{lxxi}

Nevertheless, Chavez has pledged to increase its supply of oil to the PRC significantly. In 2004, on a visit to Beijing, he told the Chinese that his country would put many of its oil facilities at the disposal of China, saying that Chinese firms would be allowed to operate 15 mature oil fields in the heart of Venezuela, capable of producing more than one billion barrels.^{lxxii} In 2008 he pledged that Venezuela would supply China with 500,000 barrels of oil per day by 2009.^{lxxiii} According to his statements, he actually came close, with 460,000 barrels.^{lxxiv} Thanks to WikiLeaks, it has been discovered that much of this oil was sold to China at as little of a price as \$5 per barrel when the market price was \$78 per barrel.^{lxxv} According to an unnamed official at PDVSA, China then sold the cheap oil elsewhere at a "sizable" profit.^{lxxvi} The Chinese willingly accept the economic concessions that a politically driven Venezuela offer them.

While the Venezuelan-produced statistics show 460,000 barrels per day in 2009, the Chinese-produced statistics show equivocating figures. For the first few months of 2010, Chinese figures showed that China was importing only an average of 132,000 barrels per day in the first few months that year. Also, a U.S. Embassy cable published by the Spanish newspaper "el País" shows mismatches on Venezuelan data, citing a PDVSA official's January 2010 estimates of oil exports to China that was nearly five times greater than the Chinese figures on such imports. To save face at not accomplishing his goal of 500,000 barrels per day, Chavez both offered oil to China at a steep discount and gave fake statistics. While the importance of bilateral trade between China and Venezuela has not reached the levels that Chavez purports it to, China does recognize the opportunities that an ideological pro-China government in Venezuela offers.^{lxxvii}

China has engaged Venezuela

successfully with bilateral deals, almost always involving petroleum. In 2007, China and Venezuela signed several bilateral agreements, the most important of which was a \$6 billion development fund to deepen collaboration in the areas of energy, technology, and financing, particularly with regards to the Sino-Venezuelan oil trade.^{lxxviii} Of the total \$6 billion, the Chinese Development Bank contributed \$4 billion, while the Venezuela's National Development Fund contributed \$2 billion.^{lxxix} By 2011, the fund has reached \$12 billion, with another \$4 billion coming from China and \$2 billion from Venezuela.^{lxxx} Also in 2007, Venezuela signed agreements with the China National Petroleum Company (CNPC) for cooperative projects in the petroleum sector, most importantly an increased participation of CNPC in Venezuela's Orinoco reserves. There have been over 300 bilateral investment commitments between the two countries, and by 2009, China became Venezuela's second-largest trading partner.^{lxxxi}

China, with vast reserves of currency, takes advantage of Venezuela's inability to attract meaningful private investment. Since private companies are often wary of investment into Venezuela, many development projects are underfunded. But with the political connections between China and Venezuela, China has more confidence in Venezuela's ability to pay back its loans in oil. The China Development Bank (CDB) provided a soft loan in April 2010 to Venezuela of \$20 billion, to be repaid over a ten year period, guaranteed by oil supplies.^{lxxxii} By trading its vast reserves of currency, China guarantees access to Venezuela's oil.

China has aggressively pursued state level deals with Venezuela for petroleum. Chávez offers China an opportunity to gain a foothold in the region- the goal of which in the end is greater energy and natural resource security. China obviously is using Venezuela's political environment and statist leader to get access to their oil and has successfully taken advantage of the political situation to do so.

Conclusion

Exports of Latin American primary products to China have increased tremendously since the turn of the 21st century. China's economic growth, powered by its high-tech exports and responsible governance, is the main reason for the uptake in these products. Latin America holds some of the world's largest reserves of minerals, agricultural products, and fuel. Latin American economies have adapted to Chinese demand through developments of industries and the growth of transportation infrastructure. China works with Latin American governments to gain access to their natural resources with a state-specific, pragmatic approach, adapting itself to the economic and political considerations of the countries in question. China's government also maintains oversight of the process of natural resource acquisition. The examples of China's advancement with Chile and Venezuela show the Chinese government's ability to work in starkly different economic, political, and ideological environments for the main purpose of the acquisition of natural resources.

Due to China's growing influence in the region, China's expansion into Latin America may threaten United States hegemony in the Western Hemisphere. The United States has traditionally been the economic and political hegemonic power in the Western Hemisphere. While the United States was busy dealing with issues in the Middle East such as the Iraq War and the War in Afghanistan, the Chinese advanced into the United States' "back yard." China now holds more economic and political sway with a few important Latin American countries. The United States must develop a comprehensive strategy to deal with the growth of Chinese influence in its hemisphere.

China's advancement into Latin America could also be seen as a form of neocolonialism. While many Latin American countries have feared economic dominance by and dependence on the powerful Western states like the United States and Western Europe, they may find that China is no different. Latin American economies are becoming even more dependent on natural resource exports because of the growth of the Chinese presence. It is particularly ironic that the Venezuelans work to decrease their dependence on the United States by increasing their dependence on China.

Although China's government is in name Communist, its foreign policy is driven by its economic markets. It is also ironic that the world's foremost Communist nation is now playing the role of a large, dominating economic power. Latin American nations should pay close attention to Chinese action in the region, and work to mitigate domination of their economies by China.

China's export driven economic growth has led it to Latin America for the acquisition of rich natural resources. Latin American countries have adapted to China's demand with economic developments for the extraction and movement of natural resources to the Pacific Ocean. China has engaged the different players in the region with thoughtful and often opportunistic strategies and has gained significant influence in the region. It will be interesting to see the future of Sino-Latin American relations given the possible implications of China's economic advancement in the region.

ⁱ Matt Ferchen. "China-Latin America Relations: Long-Term Boon or Short-Term Boom?." *The Chinese Journal of International Politics* 4, no. (2011): 55-86.

ⁱⁱ R. Evan Ellis. *China in Latin America the Whats and Wheres*. Boulder, CO: Lynne Reiner Publishers, 2009.

ⁱⁱⁱ Yuking Xing. "China's High-Tech Exports: Myth and Reality." *EAI Background Brief* 506, no. (2010): 1.

^{iv} OECD Factbook, http://www.oecd-ilibrary.org/economics/oecd-factbook_18147364 accessed April 30, 2011.

^v Jonathan Hol slag. "China's Resources and Energy Policy in Sub-Saharan Africa." *Report for the Development Committee of the European Parliament* (2007): 7.

^{vi} Guadalupe Paz. *China's Expansion into the Western Hemisphere Implications for Latin America and the United States*. Washington, D.C.: The Brookings Institute, 2008. 133.

^{vii} Paz 133

^{viii} Ibid.

^{ix} Ibid.

^x Ibid.

^{xi} Paz 134

^{xii} Ibid.

^{xiii} Hana R. Alberts, Forbes.com. January 12, 2010, <http://www.forbes.com/2010/01/12/car-sales-geely-volvo-business-autos-china.html> (accessed 3/27/2011).

^{xiv} Fred Gale. "China at a Glance A Statistical Overview of China's Food and Agriculture. *Economic Research Service/ USDA*". (2002), 1-2, www.ers.usda.gov/publications/aib775/aib775e.pdf. (accessed March 3, 2011).

^{xv} Javier Alvarez. "Stakeholder demands: Bringing together the CSR and Biodiversity Agendas in Latin America." *International Conference "Business and Biodiversity" Bonn, April 2008* (2008): 1-2.

^{xvi} Alvarez

^{xvii} Paz 133

^{xviii} Ibid.

^{xix} Rafael Valdez Mingramm. "China-Latin America Commodity Trade & Investment: Enduring Trends towards 2027...*SINOLATIN CAPITAL*". (2009), 5, (accessed April 30, 2011).

^{xx} Ferchen 61

^{xxi} Ibid.

^{xxii} Ibid.

^{xxii} i Sumeet Chugani. December 20, 2010.<http://www.miamiherald.com/2010/12/20/1981320/demand-from-china-reshaping-latin.html> (accessed March 27, 2011).

^{xxiv} Chugani

^{xxv} Tavia Grant, "Panama Canal Gets Revamped—and so does Global Trade," *The Globe and Mail*, accessed March 28, 2011, <http://professional.wsj.com/article/TPGLOB000020110326e73q0004l.html>.

^{xxvi} Grant

^{xxvii} Brett Borkan, "China 'Open' to FTA with Colombia," *Colombia Reports*, accessed March

29, 2011, <http://colombiareports.com/colombia-news/news/9461-china-is-open-to-an-fta-with-colombia-.html>.

^{xxviii} Ellis 14

^{xxix} "China Goes Outbound: the Case for Overseas M&A and Private Equity in Latin America." September 10, 2009.<http://lavca.org/2009/09/10/china-goes-outbound-the-case-for-overseas-ma-and-private-equity-in-latin-america> (accessed March 19, 2011).

^{xxx} Ellis 14

^{xxxi} Ibid.

^{xxxii} Ibid.

^{xxxiii} Ellis 15

^{xxxiv} Hol slag 8

^{xxxv} Ibid.

^{xxxvi} Hol slag 10

^{xxxvii} Hol slag 9

^{xxxviii} Sebastian Briozzo, "The Growing Importance of China-Brazil Trade," *FinanceAsia*, accessed April 30, 2011, <http://www.financeasia.com/News/239307/the-growing-importance-of-china-brazil-trade.aspx>.

^{xxxix} Ellis 34

^{xl} Ibid.

^{xli} Ibid.

^{xlii} Ibid.

^{xliii} Ibid.

^{xliv} Ibid.

^{xlv} Ibid.

^{xlii} Ellis 35

^{xlvii} "China and Chile," *Ministry of Foreign Affairs of the People's Republic of China*, accessed May 1, 2011, <http://www.mfa.gov.cn/eng/wjb/zzjg/lsmzs/gjlb/3478/t17347.htm>.

- ^{xlviii} Patricio del Sol, "Chilean Regional Strategies in Response to Economic Liberalization," *Universia Business Review*, 2010: 115.
- ^{xlix} "China, Chile, Put Free Trade Agreement into Effect," *China Economic Net*, accessed April 30, 2011, http://en.ce.cn/Business/Macro-economic/200610/02/t20061002_8814600.shtml
- ¹ Ellis 34
- ^{li} "China, Chile to Expand Trade Ties," *Wall Street Journal Industry Updates*, accessed March 29, 2011, <http://professional.wsj.com/article/TPBDU0000020100802e6820001g.html>.
- ^{lii} Ellis 37
- ^{liii} Ibid.
- ^{liv} Ibid.
- ^{lv} Ibid.
- ^{lvi} "China Trade Statistics." <http://globaledge.msu.edu/countries/china/tradestats/> (accessed March 18, 2011).
- ^{lvii} China Trade Statistics
- ^{lviii} "China, Chile to Expand Trade Ties," *Wall Street Journal Industry Updates*.
- ^{lix} "U.S. Energy Information Administration," Country Analysis Briefs: Venezuela, accessed April 30, 2011, <http://www.eia.doe.gov/countries/cab.cfm?fips=VE>, March 2012.
- ^{lx} Larry Pascal. "The Venezuelan Oil and Gas Sector – Are There Still Opportunities in the Era of Petronationalism?." *Latin American Law & Business Review* 16, no. (2008): 1.
- ^{lxi} Cezar J. Alvarez. <http://www.cfr.org/economics/venezuelas-oil-based-economy/p12089> (accessed March 27, 2011).
- ^{lxii} Alvarez
- ^{lxiii} "U.S. Energy Information Administration"
- ^{lxiv} Ellis 108
- ^{lxv} Ellis 109
- ^{lxvi} "MercoPress," May 5th, 2007.<http://en.mercopress.com/2007/05/05/only-conoco-phillips-resisting-chavez-orinoco-basin-control> (accessed April 3, 2011).
- ^{lxvii} May 5th
- ^{lxviii} "Chavez: Bush 'devil'; U.S. on its way down." September 20, 2006. http://articles.cnn.com/2006-09-20/world/chavez.un_1_venezuelan-president-hugo-chavez-nuclear-technology-veto-power?_s=PM:WORLD
- ^{lxix} Jorge Dominguez. "China's Relations With Latin America: Shared Gains, Asymmetric Hopes." *Inter American Dialogue* (2006): 42.
- ^{lxix} Alvarez
- ^{lxxi} Alvarez
- ^{lxii} "Wikileaks:China Profits off Cheap Venezuelan Oil" *Wall Street Journal China Realtime Report*. (2010), <http://blogs.wsj.com/chinarealtime/2010/12/15/wikileaks-china-profits-off-cheap-venezuelan-oil/>. (accessed March 27, 2011).
- ^{lxiii} Wikileaks
- ^{lxxiv} Ibid.
- ^{lxv} Ibid.
- ^{lxvi} Ibid.
- ^{lxvii} Ibid.
- ^{lxviii} Chris Carlson. "Venezuela and China form Bilateral Development Fund." November 7, 2007.<http://venezuelanalysis.com/news/2812>
- ^{lxix} Carlson
- ^{lxxx} Ibid.
- ^{lxxi} Anthony Bryan. "What is China doing in Latin America?Business Journal". (2010), <http://bizjournalonline.com/?p=1163..> (accessed March 27, 2011).
- ^{lxii} Simon Romero. "Chávez Says China to Lend Venezuela \$20 BillionNew York Times". (2010), <http://www.nytimes.com/2010/04/19/world/americas/19venez.html>. (accessed April 30, 2011).

Engaging the Diaspora: Prospects and Pitfalls

Janna Cisterino

Current international migration flows are quite remarkable. In 2011, the World Bank estimated that more than 215 million people, or approximately 3% of the world's population, lived outside their countries of birth.¹ There are three general types of sending states or countries that produce a significant number of emigrants: labor-exporting states, states that have gained independent statehood after a major outflow of their nationals, and states in conflict.² One of the major contributors to emigration trends in the current era is the expansion of capitalism. The real and growing demand for immigrant labor in industrialized countries has encouraged migration on a greater scale than in the past.³ Of course, other dynamics also play a role in current migration trends. There have long been linkages between emigrants and their home countries, but the scale, structure, and strength of these relationships has changed in recent years in significant ways. This paper will examine the logic behind sending states' engagement with their diasporas, the mechanisms that they employ in achieving this goal, as well as some of the potential drawbacks of engagement.

Globalization and new, improved technologies, such as in communication and travel, have altered the dynamics of transnationalism and emigrants' relationships with their sending states, making linkages between the diaspora and the homeland denser and more rapid.⁴ It is now far easier and cheaper for emigrants to call home, to visit their home countries, and to consume media produced in their home countries. These developments have strengthened the identification of emigrants with their sending states and cemented

their conception of themselves as a part of the political communities of their home countries. These transnational communities have challenged the traditional conception of national communities as defined by a common connection to some territory. As Kim Barry says, "Migration decouples citizenship and residence."⁵ Similarly, as Adamson argues, there is an "uneasy fit between the state as an administrative unit and the state as a spatially discrete homogenous political identity."⁶ Emigration disrupts the simplistic conception of the state and its nationals as a neat package bounded by the state's borders. Indeed, when emigration from a state is common, it can become a defining feature of the national experience. In the context of this paper, a diaspora will be defined as a social group of emigrants who have sustained a collective identity, maintained ties with a real or imagined homeland, and displayed an ability to address the needs of its members through organization and transnational links.⁷

Even in the context of transnational processes, politics often remains about local issues. In many cases, transnational politics is highly territorial in focus and goals, although its actors have been deterritorialized.⁸ In the past, emigrants' economic success and social status depended on their successful acculturation and integration into mainstream circles of society in their host country. However, these things now often rely upon emigrants' strong transnational links.⁹

¹ The World Bank, *Migration and Remittances Factbook 2011* (Washington, D.C.: The World Bank, 2011), ix.

² Eva Ostergaard-Nielsen, "International Migration and Sending Countries: Key Issues and Themes," in *International Migration and Sending Countries*, ed. Eva Ostergaard-Nielsen (New York: Palgrave Macmillan, 2003), 6.

³ Alejandro Portes, Luis E. Guarnizo, and Patricia Landolt, "The Study of Transnationalism: Pitfalls and Promise of an Emergent Research Field," *Ethnic and Racial Studies* 22 no. 2 (1999): 228.

⁴ Jose Itzigsohn, "Immigration and the Boundaries of Citizenship: The Institutions of Immigrants' Political Transnationalism," *International Migration Review* 34 no. 4 (2000): 1130.

⁵ Kim Barry, "Home and Away: The Construction of Citizenship in an Emigration Context," *New York University Law Review* 81 no. 11 (2006): 3.

⁶ Fiona Adamson and Madeleine Demetriou, "'Remapping the Boundaries of 'State' and 'Identity': Incorporating Diasporas into IR Theorizing,'" *European Journal of International Relations* 13 (2007): 490.

⁷ Ibid., 497.

⁸ Terrence Lyons and Peter Mandaville, "Global Migration and Transnational Politics: A Conceptual Framework," *George Mason University Project on Global Migration and Transnational Politics*, Working Paper no. 1 (2008): 6.

⁹ Portes and Guarnizo, "Emergent Research Field," 229.

The Diaspora and Economic Development

One of the reasons that sending states engage their diasporas is to obtain resources for economic development. However, there is serious disagreement about whether emigration is beneficial for a state's development. The negative perspective posits that emigration and the sending state's resultant reliance on workers' remittances results in dependency, instability, developmental distortion, and a resulting economic decline that is not worth the temporary economic gains.¹⁰ Remittances are viewed as a potentially unstable, unpredictable source of revenue for the state. A reliance on emigrants and their contributions to the state for developmental purposes sends the state down a dangerous path to political instability and unpredictable social change without the necessary institutional adjustments. The influx of money ultimately changes the state in ways that it is unprepared to constructively cope with. The positive perspective views emigration and the financial inflows that stem from it as forces for political peace and positive social changes in families and gender roles.¹¹ Although contributions from emigrants are certainly not a quick-fix, they do provide an important source of revenue to bolster economic development.

Remittances

Remittances from emigrants are now the primary source of financial inflow for many sending countries.¹² Official estimates claim that the sum of remittances is greater than overall development aid worldwide. Unofficial estimates go even further, asserting that remittance flows may be twice as large as development aid.¹³ The large growth in remittances has sharply increased developing states' economic dependence on them in the past few decades. One of the reasons for the trend in the growth in remittances is the impact of neoliberal economic policies enacted by many new democracies in the developing world. These policies have contributed to the start or

continuation of labor migration and thus the rising importance of remittances as they grow relative to other types of financial inflows.¹⁴ Further, remittances have become the top source of foreign currency for many labor-exporting countries. Finally, trends in other forms of financial inflows to developing countries have made remittances even more significant. For example, capital market flows and foreign direct investment (FDI) have fallen in the past few years, but remittance inflows continue to grow.¹⁵

The economic effects of remittances are wide-ranging. In their most obvious function the remittances that family members in the sending state receive allow them to better meet their immediate consumption needs and improve their quality of life. The importance of this function of remittances is often belittled. Admittedly, improving the standard of life of individuals in the sending country through outside contributions is less desirable than doing so by generating internal economic growth, but it is still significant. The majority of remittances that are not spent on immediate consumptions needs such as food and clothing are spent on health care and education.¹⁶ A healthier and better-educated work force is better able to contribute to economic growth.

Remittances are also important on a broader macro-economic scale. First, remittances have multiplier effects on the economy of the sending state.¹⁷ This means that a given inflow of remittances will generate a greater increase in the receiving country's gross output than the amount of remittances. By spending remittances, families contribute to the growth of their countries' economies by providing income for others and creating jobs. When the remittances that families receive are greater than their current spending needs, they raise the national savings and provide funds for loans and investment, thus generating further economic growth. Further, as mentioned earlier, remittances are an important source of foreign exchange for sending countries. This

¹⁰ Charles B. Keely and Bao Nga Tran, "Remittances from Labor Migration: Evaluations, Performance, and Implications," *International Migration Review* 23 no. 3 (Autumn 1989): 501.

¹¹ Ibid., 500.

¹² Barry, "Home and Away," 7.

¹³ Ostergaard-Nielsen, "International Migration and Sending Countries," 15.

¹⁴ OECD, "International Migration Remittances and Their Role in Development," *International Migration Outlook* (2006): 240.

¹⁵ Keely and Tran, "Remittances from Labor Migration," 504.

¹⁶ Barry, "Home and Away," 7.

foreign currency can be used to finance the country's deficits, debts, and import costs.¹⁸ Finally, remittances can provide the resources for a transition to what would otherwise be unsustainable development.¹⁹ Again, remittances in themselves cannot solve development problems, but they do provide an economic boost and can serve as a bridge to sustainable development.

Those emigrants who send remittances to their home countries do so for a variety of reasons, including pure altruism, pure self-interest, informal agreements between these individuals and their families, and portfolio management decisions.²⁰ First and most obvious is pure altruism, or a desire to either improve one's family's standard of living or to better one's home country. A second motive is pure self-interest. One case of this is when emigrants want to inherit from their older family members in the home state and expect that sending remittances will ensure that by establishing goodwill. A second example is when emigrants send remittances to ensure that the property or assets that they have left behind in the sending country are properly cared for, sometimes because they expect to return to their home country in the future. Next, there are informal agreements between emigrants and their family members in their home countries. In this case, emigrants remit as a form of insurance against financial problems in their host state, but also as a form of family loan. These informal agreements blend aspects of both the pure altruism and pure self-interest models. Finally, a fourth motive of emigrants in sending remittances to their home countries involves portfolio management decisions. For example, saving in the home country may be more profitable for emigrants. The motives behind an individual emigrant's decision to remit will not fit neatly into one of these categories. Rather, they will likely include several aspects of each.

However, there are a number of disadvantages associated with the role that remittances play. First, as mentioned earlier, some analysts argue that remittances distort the economies of sending countries. They contend that remittances are rarely used to generate capital or create jobs. Instead, they are primarily expended on consumer goods. These consumer goods have a high import content and thus reduce the country's supply of foreign

exchange.²¹ Critics of the positive impact of remittances also argue that they are a potentially instable source of revenue. However, quantitative evidence does not support the notion that there is a steep and sudden decline in remittances after a reduction in migration.²² Rather, it has been found that remittances actually increase in the wake of such a reduction after a short interval of adaptation. There are a number of reasons for this. First, a decrease in the outflow of emigrants does not decrease the current number of emigrants who are living abroad and sending remittances home. Many of these individuals remain in their host countries when overall migration decreases. Next, even when overall official estimates of migration decrease, emigrants continue to leave, both legally and illegally. It has also been found that in many cases when entire families are reunited in host countries they continue to remit to their home country regardless. Finally, when a reduction in migration results in emigrants returning home, they often bring a lump sum of earnings with them. Another argument disputing the importance of remittances is that they do not alter the fundamental weaknesses of economies. Instead, they serve as a temporary fix that masks the true economic problems. Further, remittances have failed to narrow the North-South economic gap, although they have narrowed income disparities within sending countries.

There are also problems associated with remittance-sending infrastructure. Emigrants often perceive official remittance channels as corrupt, inefficient, and unreliable.²³ Admittedly, in many cases they are. However, although using formal channels is more expensive than using informal channels in many cases, there are lower risks associated with their use.²⁴ A further obstacle is the legal and institutional impediments that often stand in the way of emigrants utilizing formal channels, such as a lack of proper identity documents.²⁵ This is particularly a problem for illegal immigrants. Sending states can take a number of steps to improve the quality of official remittance-sending infrastructure and thus encourage emigrants to utilize these formal channels. First, states must expand access to these channels to compete with

¹⁸ Keely and Tran, "Remittances from Labor Migration," 514.

¹⁹ Ibid., 503.

²⁰ Barry, "Home and Away," 10.

²¹ OECD, "International Migration Remittances," 151.

²² Ibid., 152.

more readily available informal remittance systems. Next, sending states should make efforts to reduce the transaction costs associated with these channels by combatting corruption and inefficiencies as these problems are one of the main reasons that emigrants do not utilize them. States should also refrain from interfering with private transfers by trying to force them through official channels or in to specific programs.²⁶ Coercion will only further discourage emigrants from using this infrastructure. Finally, states can try to assist emigrants whose lack of proper identity documents prevents them from using official channels. For example, Mexican consulates in the United States have begun to issue “matriculas” or identification cards for Mexican citizens living in the U.S.²⁷ However, such measures are not a full solution as matriculas are accepted by some banks for the purpose of establishing a bank account and sending remittances, but not by others.

Other Financial Contributions

Although remittances are by far the largest source of financial inflow from emigrants for sending countries, there are other important forms of financial contributions. Taxes are one other way in which sending states can extract resources from their diasporas. However, imposing a tax on the diaspora is problematic in many regards. First, states generally lack the enforcement power to coerce individuals who are not resident in their territory to pay taxes.²⁸ Additionally, many sending states do not have sufficient resources to pursue these taxes. One surprising example of the success of a tax on the diaspora is the case of Eritrea. The vast majority of the Eritrean diaspora paid a voluntary 2% income tax that was instituted after Eritrea gained its independence. This may be a function of Eritrea’s extension of dual citizenship to the diaspora and the role it accorded the diaspora in crafting the state’s constitution at independence.

Sending states can also provide incentives for emigrants to invest in their home countries. India is at the forefront of this strategy, giving select

members of its diaspora preferential treatment under investment and banking laws.²⁹ Foreign investment is an important source of foreign exchange and revenue for spurring economic growth in sending states. In addition, many sending states prefer to obtain funds from their diasporas rather than other foreign investors or international capital markets. The diaspora is often trusted more than other potential contributors by the sending state. A final important form of financial contributions is emigrants’ direct contributions for specific initiatives. Many local governments in sending states actively reach out to emigrants to obtain funding for local development projects.³⁰

Knowledge and Skill Contributions

Financial contributions are not the only way in which emigrants can assist their home countries. Contributions of human capital, such as skills and knowledge, are also important. The potential human capital contributions from diasporas are significant, especially in Africa.³¹ Such contributions are particularly salient in civil service reconstruction. The knowledge and skills of the diaspora can help to rebuild the state’s institutions, increase civil service capacity, and thus contribute to stability and peace. Diasporas are especially important in reconstruction because donors, whether individuals or other states, have short attention spans and will likely not commit to provide assistance for an extended period of time.

The motives of the diaspora in contributing human capital to their home countries are varied. One would assume that the less time an emigrant has spent away from his home country, the more likely he would be to assist in such efforts. However, it has been found that there is not necessarily a linear relationship between the time an emigrant spends away from his home country and his interest in it.³² Crises in the homeland are one important factor in sparking emigrants’ interest in and human capital contributions to their home

²⁶ Alan Gamlen, “Diaspora Engagement Policies: What are They, and What Kinds of States Use Them?,” *Centre on Migration, Policy, and Society*, Working Paper no. 32 (2006): 15.

²⁷ Susan Sachs, “New York, Citing Security, Rejects Mexican ID Cards,” *New York Times*, December 28, 2002.

²⁸ Barry, “Home and Away,” 10.

²⁹ Ibid., 11.

³⁰ Itzigsohn and Villacres, “Migrant Political Transnationalism,” 676.

³¹ Jennifer M. Brinkerhoff, “Exploring the Role of Diasporas in Rebuilding Governance in Post-Conflict Societies,” in *Africa’s Finances: The Contribution of Remittances*, New Castle upon Tyne, (UK: Cambridge Scholars Publishing, 2008), 242.

³² Ibid., 243.

country. Emigrants' desire to express their identity is also a causal factor in this case, as was seen with other types of contributions. Finally, in a more pragmatic sense, emigrants often become involved in human capital transfers to maintain or acquire resources.³³

Knowledge transfers are especially pertinent in rebuilding governance in post-conflict societies. There are three dimensions to post-conflict reconstruction: a rehabilitative dimension oriented to the past, a resolute dimension oriented to the present, and a preventive dimension oriented to both the present and the future.³⁴ Good post-conflict governance needs to ensure effectiveness, legitimacy, and security. External actors, including diasporas, can influence reconstruction and prevent the reemergence of conflict by contributing resources, aiding in socialization, and providing political incentives for stability.

There are a number of specific programs that harness the human capital of emigrants to assist sending states. The United Nations Development Programme (UNDP) administers the Transfer of Knowledge Through Expatriate Nationals (TOKTEN) program, which focuses on the role that emigrants' human capital can play in economic development in sending states.³⁵ Other related programs include the International Organization for Migration's (IOM) Return of Qualified African Nationals (RQAN) and Migration for Development in Africa (MIDA) programs and the Afghanistan Expatriate Program (AEP) organized by the Afghan government and donors. These programs are designed to combat the brain drain in developing countries and gather the knowledge and skills necessary for development. There are a number of forms of knowledge exchange that these programs can focus on, including permanent and temporary emigrant return. However, a number of studies have found that permanent return programs are not very cost-effective and are thus less desirable to resource-strapped states. In addition, recruitment for programs that involve emigrants' return, whether temporary or permanent, is dependent on emigrants' secure legal status in their host states. An emigrant is unlikely to take part in a program that could endanger his chance of remaining in his host country.

There are a number of steps that sending states can take to increase emigrant participation in

knowledge transfer program. First, states must be clear about the program's objectives, both to the emigrants and to themselves. This is essential to effectively plan recruitment strategies and develop participant selection criteria. Next, there should be clarity about the duration of the program and emphasis should be placed on the specific skills and credentials that will best contribute to the program's success. The incentives and compensation provided to emigrants should be targeted, have minimal visible perks, and be based on a market analysis.³⁶ The problems associated with excessive compensation will be discussed later. Next, the sending state and any international organization that it is working with should identify and utilize existing diaspora organizations and networks for recruitment and program support. Emigrants who participate in these groups are more likely to maintain interest in their home country and be willing to contribute human capital through knowledge transfer programs. Finally, and perhaps most importantly, these programs should be conducted under the authority of the sending state's government. Attempts to bypass the state by other actors, perhaps with good intentions, only serve to weaken the state's legitimacy and create governance problems in the future. Any other actors, such as international organizations, that are involved should not try to co-opt the sending state's authority.

Yet, there are a number of potential problems associated with knowledge transfer programs. As already mentioned, bypassing nascent or weak governments with NGOs, IGOs, and emigrants can undermine the governments' authority, and it can also drain local capacity and entail greater costs. Employing emigrants, even under the auspices of the government, can also be problematic. First, the repatriation of the diaspora can lead to the emergence of a new political elite and create or exacerbate political tensions. This elite is likely to be out of touch with their home country to some extent due to the time they have spent away. Further, it is hard to recruit members of the diaspora with appropriate skills for these projects. Those with talent are generally among the first to leave the country, and have often been away for a significant period of time. In addition, they are the most likely to successfully integrate into their receiving societies, and thus are less inclined to return even for short-term projects. Those who are recruited must often be offered compensation at

³³ Ibid., 243-4.

³⁴ Ibid., 245.

³⁵ Gamlen, "Diaspora Engagement Policies," 16.

³⁶ Brinkerhoff, "Exploring the Role," 257.

international wage levels and may be resented by the less well-paid indigenous civil service.³⁷

Political Advocacy

Diasporas can also be influential in political advocacy on a number of issues. The motives of the diaspora in advocacy are generally quite similar to those for maintaining ties with the sending state in general. These motives include acquiring power and resources, assuaging guilt about leaving the sending state behind, maintaining a collective memory or myth of their home country, expressing an ethnic group consciousness, and keeping alive the expectation of return, among others. An important addition to the motives in the case of advocacy is a desire to improve bilateral relations between the emigrants' sending and receiving states. The most prominent examples of diaspora advocacy are the Israeli and Irish diasporas. The Israeli diaspora has been instrumental in lobbying their host governments to obtain support for Israel, and the Irish diaspora has been instrumental in fund-raising and mustering political support for a united, independent Ireland.³⁸

Effective diaspora advocacy hinges on a number of factors. First, the diaspora must have adequate resources and the means to use them.³⁹ In other words, they must have a supply of money and the organizations and mechanisms to use this money to achieve a desired political outcome. The sending state's support of a diaspora is crucial in establishing legitimacy for the group and assisting in the coordination of advocacy efforts. The target country's political system is also a factor. Democratic, liberal governments are far more open to emigrants' lobbying efforts than other types of governments. In recent years, democratization in many sending countries has opened a new avenue for diaspora political participation. New technologies that facilitate transnational politics are also being seized upon by both non-state political entrepreneurs and state elites who are using the organizational form of "diaspora" and diaspora mobilization as a means of generating material resources and political support.⁴⁰

³⁷ Ibid., 248-9.

³⁸ Ostergaard-Nielsen, "International Migration and Sending Countries," 18.

³⁹ Newland, "Voice After Exit," 2.

⁴⁰ Adamson and Demetriou, "Remapping the Boundaries," 491.

Diasporas are concerned about a number of issues and pressure different actors to achieve their desired political outcomes. First, diasporas care about their status in both their sending and host states. They would lobby both states on these issues. Their status is regulated in such matters as citizenship, migration status, voting rights, and their ability to participate in the socioeconomic lives of their sending towns or regions. A second area of diaspora interest is issues that affect their sending countries. This sector includes concerns such as human rights, good governance, and political participation in the homeland, both for themselves and for their fellow nationals. Finally, there are issues with bilateral impact, including trade policy, humanitarian relief, development policy, diplomatic recognition, and economic sanctions.⁴¹ The political character of the sending country shapes the issues that the diaspora focuses on. In countries where democracy is just emerging, emigrants usually focus on local politics and development. On the other hand, in countries where there is a high degree of political institutionalization and partisanship, emigrants are more likely to demand that they be allowed to participate in the political process.⁴²

Diaspora advocacy takes on a number of forms. Emigrants can function as lobbyists, putting direct pressure on a government to change its policies. Other forms of direct political participation include electoral participation and public demonstrations. Emigrants may also initiate lawsuits to elicit a change in government policy or conduct media campaigns to alter public opinion. Finally, emigrants can support political parties or causes through fund-raising.⁴³ Diasporas' interaction with political parties has taken on new forms in the age of transnational politics. The inclusion of political organization organized abroad in the democratic political system of the sending states is new and quite remarkable.⁴⁴ Most diaspora advocacy is directed toward either their sending or receiving countries. However, actors such as NGOs, international organizations, media, business, labor unions and churches are also targets of diaspora action.⁴⁵

However, as with every other aspect of diaspora engagement discussed thus far, there are

⁴¹ Leblang, "Harnessing the Diaspora," 8.

⁴² Itzigsohn, "Boundaries of Citizenship," 1146.

⁴³ Newland, "Voice After Exit," 7.

⁴⁴ Itzigsohn, "Boundaries of Citizenship," 1144.

⁴⁵ Newland, "Voice After Exit," 7.

potential drawbacks and difficulties associated with diaspora advocacy. First, there is the issue of political fragmentation within the diaspora and thus considerable difficulty in ascertaining who can legitimately speak for the diaspora. Mexico has addressed this problem through the Consultative Council of the Institute for Mexicans Abroad, which consists of elected leaders from diaspora communities.⁴⁶ With this institution, Mexico has been able to conduct relations with members of the diaspora who are at least somewhat representative of the diaspora's views. A second issue is when a diaspora's advocacy is undermining, not supporting, a sending state's position. One notable example of this is the case of Turkey and its Kurdish diaspora. The Kurdish diaspora has been largely antagonistic towards Turkey and has challenged the official conception of Turkish nationalism. In host countries across Europe, the Kurdish diaspora has been able to interact and build networks that bypass, challenge, and contest identity construction in Turkey.⁴⁷ Situations such as this pose serious problems for sending countries, as they do not have the authority or capacity to silence dissent abroad.

Dual Citizenship/Nationality and Voting Rights

One of the ways in which sending states can engage their diaspora populations is by extending voting rights or dual citizenship to these individuals. In this section, the extension of dual citizenship to the diaspora and allowances for external voting will be treated together. However, one should note the difference between the two. External voting simply refers to allowing citizens to cast votes outside the sending state. If a sending state does not allow dual citizenship, emigrants who have acquired the citizenship of their host country will not be able to participate. In addition, a sending state that allows dual citizenship may not have provisions for external voting. In this case, an emigrant who wants to vote in his sending country's elections must return home.

Although many states perceive the phenomenon of external voting as anomalous and politically unpalatable, it has in fact become quite common in recent years. Various sources provide slightly different data on external participation in elections, but all point to a growing voice for

emigrants in home country politics. Of the 144 countries on which data is available, 80% allowed emigrant participation in elections in some form.⁴⁸ According to David Leblang, 91 of 186 countries allow expatriates to vote, and 84 countries permit dual citizenship.⁴⁹ The most prevalent form of external voting is voting abroad for one's home district. This method compromises 61.8% of the sample of external voting in Collyer and Vathi's study.⁵⁰ It has also been found that democracies are more likely to extend these benefits to their diasporas than non-democracies. However, some non-democracies have also accorded their diasporas external voting rights. For example, emigrants were able to participate in the 2007 Moroccan elections.⁵¹

However, the concept of dual citizenship is convoluted. There are two common, and quite dissimilar, interpretations of this term.⁵² The first conception of dual citizenship is one that conforms to the everyday understanding of the term. In this conception, when a state extends dual citizenship to members of the diaspora this includes the full rights accorded to citizens resident in the state, including voting. However, there is another interpretation of the term. The extension of dual citizenship can also serve as a more symbolic gesture. When states adhere to this interpretation, emigrants are recognized as members of the national community, but they are not granted all of the rights and benefits of traditional citizenship, such as voting and holding office.⁵³ This restricted dual citizenship is sometimes termed dual nationality, though this is far from universal practice. One must be careful in distinguishing between those states who allow full political participation for non-resident citizens and those who do not as the term dual citizenship is used with a consistent meaning in neither academic nor policy publications.

Diasporas are interested in and feel entitled to dual citizenship for a number of reasons. On an emotional or psychological level, emigrants desire

⁴⁸ Michael Collyer and Zana Vathi, "Patterns of Extra-territorial Voting," *Sussex Centre for Migration Research Working Paper T22* (October 2007): 15.

⁴⁹ David Leblang, "Harnessing the Diaspora: The Political Economy of Dual Citizenship," *University of Virginia* (February 2010): 3.

⁵⁰ Collyer and Vathi, "Patterns of Extra-territorial Voting," 15.

⁵¹ Ibid., 5.

⁵² Barry, "Home and Away," 4.

⁵³ Ostergaard-Nielsen, "International Migration and Sending Countries," 19.

⁴⁶ Ibid., 4.

⁴⁷ Adamson and Demetriou, "Remapping the Boundaries," 509.

to conceive of themselves as parts of the state whose territory they have left behind. They want to be recognized and legitimated by their sending states' governments as being a part of their respective communities. Second, many emigrants believe that they still have a stake in their home countries. Improved communication and transportation technologies allow emigrants to remain in contact with family members in the sending state, to send remittances, to visit, to retain property and invest in the sending state, and to harbor hopes of returning one day. These linkages make emigrants feel that they are still an integral part of their home countries.⁵⁴ Finally, voting and other non-resident citizenship rights provide a way for emigrants to maintain and develop legitimate, state-sanctioned forms of political transnationalism.⁵⁵ Although they may not utilize them in every case, emigrants desire access to formal transnational channels. Again, this is a function of their desire for recognition, the need to be treated as legitimate actors.

The main interest for sending states in extending dual citizenship is the expectation that they will be able to expect more from their diasporas in return. Sending states hope that the extension of dual citizenship will cement emigrants' identification with their home countries and thus elicit increased contributions. They hope that according rights to emigrants will instill them with a sense of obligation. In addition, extending political rights may foster a sense of trust between the sending state and the diaspora. This is crucial in establishing and maintaining productive working relationships. These relationships can channel increased contributions, whether they be financial or of a different nature, to the sending state. The extension of dual citizenship may increase the financial inflow from emigrants, but can also make emigrants more likely to return home, a contribution of human capital.⁵⁶ Another rationale behind extending voting rights to the diaspora is specifically related to post-conflict societies. In countries where conflict has intentionally resulted in the expulsion of a particular ethnic or other social group, it is crucial that external voting be allowed. When this is not the case, the state

legitimizes ethnic cleansing.⁵⁷ If external voters are excluded, the electorate and thus the resulting elected government will be heavily biased towards the social group that conducted the ethnic cleansing. States must reject the idea that a segment of their population can be removed from political participation by the use of force and their expulsion from the state.

There is quantitative evidence that there is a positive and significant correlation between the extension of dual citizenship to emigrants and financial inflows into the sending state. Leblang's study estimates that India, which does not currently permit dual citizenship or external voting (except in very limited circumstances), would likely increase the remittances that it receives from its diaspora by between 2% and 2.5% if it extends dual citizenship.⁵⁸ Other forms of financial inflows are affected as well. For example, both the migrant stock (the number of emigrants from a particular sending state that a host state has) and dual citizenship have positive and statistically significant effects on the allocation of foreign economic assistance to the sending state. This also holds true with respect to the cross-border flow of portfolio investment.⁵⁹ It seems that many states recognize the financial benefits of extending dual citizenship to their diasporas, as when a state's national debt increases by 1%, the likelihood that it will extend dual citizenship to its diaspora increases by .15%.⁶⁰ Although the strength of this relationship is far from overwhelming, it demonstrates sending states' recognition of the role their diasporas' contributions can play in ensuring their financial well-being. However, in the case of the relationship between the extension of dual citizenship and emigrants' intention to return to their home states, only a marginal effect has been documented.⁶¹

There are a number of ways in which the voting process for the diaspora can be conducted. The easiest and least costly method as far as the sending state is concerned is to only allow emigrants to vote if they return in person to their home district. However, this imposes a heavy burden on emigrants who desire to vote. The second mode of participation is voting abroad for

⁵⁴ Ibid.

⁵⁵ Collyer and Vathi, "Patterns of Extra-territorial Voting," 22.

⁵⁶ Leblang, "Harnessing the Diaspora," 4.

⁵⁷ Collyer and Vathi, "Patterns of Extra-territorial Voting," 8.

⁵⁸ Leblang, "Harnessing the Diaspora," 19.

⁵⁹ Ibid., 13-15.

⁶⁰ Ibid., 27.

⁶¹ Ibid., 22.

one's home district. In this method, emigrants are able to vote in their host countries, whether at their countries' embassies or polling stations established specifically for the occasion, and have their votes counted as if they were still resident in their home district in the sending country. The final type of electoral participation is voting abroad for one's own dedicated representatives. Voting abroad for dedicated diaspora representatives has largely been a post-1995 development. A final distinction is that in some cases the diaspora is only allowed to participate in certain types of elections, such as legislative or presidential.

However, there are a number of concerns associated with the extension of dual citizenship or providing external voting to the diaspora. First, there is the principle of one man-one vote or the idea, in this context, that an individual should not be able to voice their opinion in elections in multiple countries and thus have more influence than others. When members of the diaspora are able to vote both in their host and home states, this principle is violated. Sending states do not confront this issue when they don't permit dual citizenship but allow external voting for their emigrants who have not acquired another citizenship. A second point of contention is whether the diaspora should be able to participate in the elections of a polity whose laws they are no longer subject to. The argument here is that individuals do not have the right to decide how others will be ruled when they will not have to deal with the consequences of their electoral decisions as well. A related issue is whether diasporas are too far out of touch with their sending states' politics to effectively and intelligently participate in elections. The time they have spent away from their home countries may make them ignorant of current concerns and needs.

Next, diaspora political participation can contribute to ethnic and religious tension in the sending state in some cases. For example, although India has not extended voting rights to its diaspora, it has encountered problems with Hindutva, or Hindu nationalist, organizations that contribute money to Hindu rightist parties in India and try to influence Indian politics in other ways. Hindutva ideology and organizations have become increasingly visible in many South Asian diaspora communities since the early 1990s.⁶² For example,

⁶² Catarina Kinwall and Ted Svensson, "Hindu Nationalism, Diaspora Politics, and Nation-Building in India,"

the Vishwa Hindu Parishad (World Hindu Congress), a Hindu Indian diaspora organization, contributed to ethnic and religious violence in the 2002 Gujarat riots.⁶³ If voting rights were extended to the Indian diaspora, it is likely that Hindutva political parties would receive an electoral boost. In general, individuals who are not subject to the consequences of their electoral choices are more likely to engage in spiteful, destructive ethnic politics. The diaspora may have a romanticized conception of their homeland and may hold on to past grudges to maintain group solidarity. Alarmingly, Paul Collier has found that a country that has recently ended a conflict and has a large diaspora population is far more likely to fall into conflict again than a country that does not.⁶⁴ Finally, the extension of dual citizenship can provoke controversy in the host country.⁶⁵ It can exacerbate worries in the host state about the diaspora's loyalty. However, this proves less controversial when voting rights are not included in the model of dual nationality rather than dual citizenship. All of these concerns and others make dual citizenship far more controversial than other types of emigrant programs, such as cultural, social, or economic initiatives.⁶⁶

Diaspora Engagement Policy

In the past 5 years, diasporas' engagement with their sending countries has increased dramatically. This is a result of emigrants being better organized in their host countries, the development of new communication and transportation technologies, and sending states' recognition of the importance of engaging with their diasporas.⁶⁷ States enacting diaspora

Australian Journal of International Affairs 64 no. 3 (2010): 284.

⁶³ Kathleen Newland, "Voice After Exit: Diaspora Advocacy," *Migration Policy Institute* (2010): 16.

⁶⁴ Paul Collier, "Economic Causes of Civil Conflict and Their Implications for Policy," in *Leashing the Dogs of War: Conflict Management in a Divided World*, ed. Chester A. Crocker et. al. (Washington D.C.: United States Institute of Peace Press, 2007), 210.

⁶⁵ Barry, "Home and Away," 12.

⁶⁶ Collyer and Vathi, "Patterns of Extra-territorial Voting," 4.

⁶⁷ African Diaspora Policy Center "Building Institutional Cooperation Between the Diaspora and Homeland Governments in Africa: The Cases of Ghana, Nigeria, Germany, USA, and the UK," *African Diaspora Policy Centre Policy Brief* (June 2010): 3.

engagement policies come from all geo-political regions.⁶⁸ They are also of various development levels and have both civic and identity-based models of citizenship. Emigrant attitudes to their sending states are extremely varied and their levels of participation in transnational politics vary widely.⁶⁹ Sending states' policies toward their diasporas vary as well.

For the most part, sending states continue to conceive of their diasporas as primarily economic resources based on their contributions through remittances, investment, the consumption of the sending states' products, and their entrepreneurship.⁷⁰ Although only a tiny percentage of this financial inflow is sent directly to the government, sending states realize that it is beneficial nonetheless. States also try to promote the upward mobility of emigrants, perhaps out of pure altruism, but more likely to increase emigrants' income and increase the likelihood that they will contribute to their home countries. Emigrants who were driven out by conflict often don't have much to contribute to their home countries, but in some cases, such as with Bosnian and Eritrean refugees, these diasporas contribute a significant amount of money.⁷¹ Sending states also want to mobilize political support and control subversive political dissidence in the diaspora.⁷²

Many of the current relationships between emigrants and sending states are informal, ad hoc, and sporadic.⁷³ They are based primarily on individual and group interactions and directed specifically at the family, village, or local level. Many countries lack an official diaspora policy. This lack of official policies on emigrants hampers intensive and sustained cooperation between sending states and their diasporas, especially with regard to issues of development. Formal, institutionalized relationships at the national level between sending states and their diasporas would be beneficial in several ways. Such cooperation would lend emigrants credibility, give them a political voice, and increase their contributions to their sending countries.⁷⁴ Formal relationships would aid in the development of better diaspora and

development-related institutions as well.

Additionally, sending states may be able to contain, manage and co-opt potentially dangerous groups, such as ethnic nationalist or separatist organizations, through engagement. For example, many Indian emigrants have engaged in long-distance nationalism, including fund-raising and charity for the homeland. Much of this money has been collected by Hindu nationalist organizations, often without the donors' knowledge.⁷⁵ Instead of ignoring the problem or trying to marginalize these groups, the Indian government should try to engage them and redirect the money in more beneficial directions.

Formal engagement can also be useful in linking diaspora organizations with other pertinent actors. In general, emigrant organizations are more helpful in development work when partnered with sub-national development actors, such as NGOs, local governments in the sending country, or the private sector.⁷⁶ In these situations, the interest and resources of emigrants can be shaped by the development experience of other actors who have more technical expertise in the area. Without proper guidance, diaspora organizations may direct their energies and resources in inefficient ways.

Alan Gamlen posits that there are three different types of relationships between sending states and their diasporas: relations of power, relationships of communication, and finalized activities.⁷⁷ Relations of power concern the relative control that these actors can exercise upon each other. For example, some states have the ability to exercise a great deal more influence on their diasporas than other states do. When states hold the upper hand in their power relationships with their diasporas, they are better able to structure their interactions in ways that benefit the state. In cases where diasporas exercise a considerable amount of power, perhaps more than their sending states, they are able to direct their relationship in ways that reflect their interests most effectively. However, the outcomes of both of these types of power relationships are dependent on states' and diasporas' perceptions of their interests. If these actors perceive their interests similarly, there will be only be small disparities in outcomes based on the power distribution. But, if their perceptions of their interests are widely divergent, power relationships are instrumental in determining

⁶⁸ Gamlen, "Diaspora Engagement Policies," 20.

⁶⁹ Barry, "Home and Away," 8.

⁷⁰ Leblang, "Harnessing the Diaspora," 2.

⁷¹ Ostergaard-Nielsen, "International Migration and Sending Countries," 8.

⁷² Ibid., 4.

⁷³ ADPC, "Building Institutional Cooperation," 4.

⁷⁴ Ibid., 5.

⁷⁵ Kinwall and Svensson, "Hindu Nationalism," 285-6.

⁷⁶ ADPC, "Building Institutional Cooperation," 7.

⁷⁷ Gamlen, "Diaspora Engagement Policies," 5.

outcomes. Relationships of communication are simply the channels that sending states and emigrants utilize to maintain contact with and knowledge of each other. Finalized activities include the extension of rights to and the extraction of resources from the diaspora. Although sending states may feel threatened by the extension of rights, this can prove to be a necessary component of transnational sovereignty in cases where the diaspora demands them.

Similarly, there are also three different types of diaspora engagement policy: capacity building, extending rights to the diaspora, and extracting resources from the diaspora.⁷⁸ Capacity building has two key components: fostering diaspora identification and solidarity with the homeland and crafting appropriate governmental structures to engage and effectively use the diaspora. To encourage emigrants to identify as part of the imagined community of their home country, sending states can invoke symbols, espouse a rhetoric of national solidarity, or try to influence expatriate-focused media to carry their message. States can also support institutions such as culture and language clubs in emigrants' host countries. The state's aim in this aspect of capacity building is to create a more homogenous, and thus more easily manageable diaspora population.⁷⁹

The development of effective governmental apparatuses to deal with the diaspora requires a number of additional actions after the diaspora's identification with the homeland has been solidified. The first step generally involves the collection of basic information on the diaspora, such as data on emigrant populations or existing diaspora organizations in various host countries. With these data in hand, states can assess the methods they can use to manipulate their diasporas. One effective way for states to co-opt and manage their diasporas is creating a transnational migration organization under the authority of the state and treating it as a consultative organization.⁸⁰ Representation of the diaspora in this manner may allay some of the concerns of the diaspora in whether the state is considering its interests. This approach holds promise for states that wish to forgo extending a more formal voice to their diasporas by granting them voting rights, which can entail a number of previously mentioned problems. The process of extending rights to the diaspora was

discussed earlier, as was the extraction of resources. Obtaining resources from the diaspora is a matter of persuading emigrants and convincing them that their interests include supporting their homelands.

Some of the specific ways that sending states can engage their diasporas have been discussed earlier. To begin a conversation with their emigrant populations, states can reach out to them through existing governmental structures such as embassies or consulates. Diaspora organizations such as hometown associations also provide a forum for states to engage these populations. Besides engaging emigrants through specific institutions, sending states can take steps to encourage the diaspora to harbor favorable perceptions of their homeland. For example, states can extend political rights, encourage financial contributions through rhetorical appeals and financial incentives, and assert emigrants' rights abroad. Advocacy by sending states for their diasporas may be especially helpful in convincing emigrants that their home states are supporting them and thus are worthy of reciprocal assistance.

Mexico

The successes and failures of sending countries in engaging their diasporas are instructive in which policies are most likely to be effective and useful for similar states. Mexico is a classic example of an emigration or sending state. Over the years, it has had a consistent and significant net outflow of people.⁸¹ The majority of Mexican emigrants reside in the United States. In 2000, Mexico's National Population Council estimated that approximately 8,173,689 Mexican-born individuals resided in the United States. Such a large emigrant population centered in one host country provides a concentrated and significant financial flow for Mexico. In 2001, the Bank of Mexico reported that Mexican emigrants in the United States remitted a total of US\$8.895B. In the same year, the Inter-American Development Bank estimated this flow of remittances at US\$9.23B.⁸² Remittances constitute the third largest source of

⁷⁸ Barry, "Home and Away," 2.

⁷⁹ Ibid., 7.

⁸⁰ Ibid., 8.

⁸¹ Jesus Martinez-Saldana, "Los Olvidados Become Heroes: The Evolution of Mexico's Policies Towards Citizens Abroad," in *International Migration and Sending Countries: Perceptions, Policies, and Transnational Relations*, ed. Eva Ostergaard-Nielsen (New York: Palgrave Macmillan, 2003), 33.

foreign capital for Mexico, falling behind only oil exports and foreign investment.

Mexico's relationship with its diaspora has not always been positive or supportive. In fact, there is a past tradition of neglect of and disdain for emigrants among Mexican government officials as well as intellectuals.⁸³ Mexican emigrants were often treated as traitors to their homeland and their identity, having abandoned both to pursue a new life abroad. However, this dynamic began to change during the 1980s. In this decade, Mexican President Carlos Salinas de Gortari courted Mexican-American leaders and their organizations in search of political backing and financial contributions. From this point onward, Mexican emigrants and their organizations began to exert influence on both the Party of the Democratic Revolution (PRD) and the Institutional Revolutionary Party (PRI). In addition, the revolutionary left-wing Zapatista Army of National Liberation (EZLN) began to vie for emigrants' loyalty in 1994.⁸⁴

The increased influence of the diaspora on Mexican domestic politics emboldened emigrants to demand more representation and rights from the Mexican government. During the 1990s the Mexican government granted emigrants significant new political rights. In 1996 Mexico launched external voting for its citizens who resided abroad. Additionally, Mexico amended its constitution to permit the non-loss of Mexican nationality in the case of a Mexican acquiring a different citizenship.⁸⁵ These are two distinct developments and neither one is quite as beneficial for Mexican emigrants as one would first think.

First, the adoption of external voting did indeed allow for Mexicans abroad to cast their votes in the 2006 presidential elections by post.⁸⁶ Previously, Mexican citizens would have to return to their home district in Mexico to cast their ballots. The number of citizens who returned home to vote in the past was small but meaningful.⁸⁷ However, emigrants needed a special identification card to vote abroad. This card was issued only in Mexico, so emigrants would in fact have to return in order to

vote if they did not already possess this identification card.⁸⁸ It is estimated that 4.2 of the 11 million Mexicans abroad possessed one of these ID cards in 2006, but only 40,665 of these individuals registered to vote.⁸⁹ Second, the reform that allowed for dual nationality did not permit dual citizenship in the more permissive sense and its attendant voting rights. Rather, those Mexicans abroad who had acquired a different citizenship were allowed to regain Mexican nationality and a few accompanying rights, including certain benefits associated with land ownership and inheritance rights in Mexico. This was primarily a symbolic move that recognized these individuals as members of the Mexican national community.

These reforms were significant however and signaled a meaningful shift in the Mexican government's treatment of the diaspora. Mexican President Vicente Fox in particular recast the image of Mexican emigrants as a welcome and integral part of the Mexican community.⁹⁰ This recognition was a function of both emigrants' demands for rights and recognition as well as the government's acceptance of Mexico's increasing dependence on the diaspora's remittances, investments, and transfers of knowledge, skills, and technology. It will take several years to determine if the actions that Mexico has taken in recent years to draw its diaspora into the fold will substantially increase emigrants' contributions.

Nigeria

Nigeria is also a country that has a history of shunning its diaspora. The country has failed to see its emigrants as potential contributors to Nigeria and its development. Rather, it has characterized the diaspora, as well as internal migrants, as a sign of development failure, not a source of positive political and social change. Generally, the only connection between migration and development in Nigeria that has been acknowledged by the Nigerian government is the knowledge and skills that emigrants returning to Nigeria bring back with them.⁹¹ This negative perception of emigrants is

⁸³ Ibid., 44.

⁸⁴ Ibid., 47-8.

⁸⁵ Barry, "Home and Away," 12.

⁸⁶ "Voting from Abroad: The International IDEA Handbook," *International Institute for Democracy and Electoral Assistance* (November 2009): 17 and 23.

⁸⁷ Barry, "Home and Away," 16.

⁸⁸ IDEA, 22.

⁸⁹ IDEA, 32.

⁹⁰ Martinez-Saldana, "Los Olvidados," 52.

⁹¹ Hein de Haas, "International Migration and National Development: Viewpoints and Policy Initiatives in Countries of Origin: The Case of Nigeria" *International Migration Institute: University of Oxford* (June 2006): 14.

largely a function of Nigeria's oil supply and its consequences. Emigration of the Nigerian elite has long been associated with the outflow of capital from Nigeria and the transfer of oil profits to overseas accounts.⁹² As a consequence, emigrants have been viewed as individuals removing wealth from the country rather than contributing to it. Beyond being economic leeches, emigrants have also been characterized as social deviants and traitors.

However, the emigration of individuals from lower socioeconomics classes of Nigerian society has in fact been financially beneficial for the state. Even human trafficking, a serious problem in Nigeria, has appreciably increased remittances.⁹³ Unfortunately, Nigeria has not yet pursued a migration or remittance-focused development strategy, most likely because it does not hold emigrants in high regard. Emigrants' contributions are no longer a resource that Nigeria can ignore, as countries with far fewer emigrants have begun to examine and implement policies to harness their diasporas' power. Nigeria must shift its current focus on migration prevention to a more realistic and useful orientation.

The framework for a transition to a more effective diaspora engagement strategy may have been laid in the late 1990s. Nigeria's 1999 transition to democracy may bode well for future policy related to emigrants. As mentioned earlier, democracies are more likely than non-democracies to extend political rights to their diasporas. From this fact, one can generalize that democracies are more likely to recognize and support their diasporas. This seems to be the way Nigerian emigrants felt at the inception of democracy, as they have been more willing to return to the country as well as contribute financially.

Since its transition to democracy, the Nigerian government has taken a number of steps to engage the diaspora. For example, a "Presidential dialogue with Nigerians abroad" was launched in 2002, and the government has established the Nigerian in the Diaspora Organization (NIDO) in the Ministry of Foreign Affairs.⁹⁴ This organization's official mission is four-fold. First, it aims to increase the participation of the diaspora in Nigerian affairs. Second, it serves as a communication forum between the state and the diaspora. Third, it is intended to improve Nigeria's international image.

⁹² Ibid.

⁹³ Ibid.

⁹⁴ Ibid., 17.

Finally, the organization is focused on building a network of Nigerians with professional skills and knowledge and making this data available to the Nigerian government, the private sector, and other actors who partner with Nigeria on development issues.

However, Nigeria's perception of migration as only being beneficial when emigrants return is still evident. The recently created Nigeria National Volunteer Service (NNVS) was crafted to reverse the effects of the brain drain.⁹⁵ This reflects the emphasis placed on the value of return migration. This program has only been able to achieve the temporary return of skilled emigrants in most cases. Nigeria has also been largely dismissive of remittances because it has focused instead on larger transfers of funds, generally linked to the oil industry. Oil revenues have made remittances less significant as a percentage of Nigeria's GDP.

To harness the power of its diaspora, Nigeria must revamp its anti-emigrant image and sustain efforts to communicate and create relationships with the diaspora. The government must address the exploitation of Nigerians abroad, which has become a considerable problem, especially with respect to human trafficking.⁹⁶ Advocacy for the human rights of emigrants will increase trust between the government and the diaspora. Beyond abandoning its anti-emigration stance, the Nigerian government should embrace more liberal migration policies. Many of its current policies are focused on restricting the free movement of individuals. Liberalizing migration policies may promote development, as individuals who are freely able to travel back and forth to a country are more likely to view the country favorably and contribute to it.⁹⁷ There are a number of other unfavorable structural conditions that need to be remedied. High crime rates, governmental corruption, an unfavorable investment environment, and general insecurity have discouraged emigrants from investing in or returning to Nigeria.⁹⁸

The Dominican Republic

The case of the Dominican Republic and its dealings with its diaspora reflect the dynamics of diaspora lobbying of the homeland, the institutionalization of democracy and competitive

⁹⁵ Ibid.

⁹⁶ Ibid., 21.

⁹⁷ Ibid., 24.

⁹⁸ Ibid., 23.

party politics, and the extension of political rights to the diaspora. Dominican emigrants have been pressuring their home government for recognition and political rights since the 1970s. However, until the mid-1990s the Dominican government largely ignored its diaspora beyond accepting the money that these individuals remitted. In 1994, the Dominican Constitution was amended to permit Dominican citizens resident abroad to acquire another citizenship without losing their Dominican one.⁹⁹ This was a step further than the Mexican government went in its non-loss of Mexican nationality reform.

The diaspora became even more entwined in Dominican politics in 1996, the year in which competitive multi-party elections were institutionalized in the DR. In 1996, Leonel Fernandez, who spent most of his childhood in New York as a legal permanent resident, was elected President, reflecting the growing influence of Dominican emigrants.¹⁰⁰ In 1997, the right to vote abroad was extended, but was not implemented until the 2004 presidential elections. Citizens resident abroad may vote in presidential elections and are only permitted to cast ballots in person from a select group of cities in five countries: the United States, Canada, Spain, Puerto Rico and Venezuela.¹⁰¹ The diaspora electorate is particularly significant in the case of the Dominican Republic, as the emigrant population eligible to vote represents approximately 25% of the 5 million voters registered in the country.¹⁰² In 2004, of the 52,431 Dominican citizens registered abroad, 35,042 cast ballots in the presidential election, showing an impressive turn-out rate of over 66%.¹⁰³ Although the turn-out for registered voters abroad was quite high, the vote abroad reflected only about 1% of the overall vote in this election. A conservative estimate would predict this number to be perhaps 10 times greater.¹⁰⁴ Perhaps Dominican emigrants simply wanted to be recognized by their home government and did not have a strong desire to actually participate in the DR's elections. Another possible explanation is the requirement that Dominican voters abroad possess an

identification document that is only available in the DR, similar to Mexico's requirement, discouraged many potential voters. Emigrants who were not fully committed to participating in the Dominican electoral process or ones whose economic means made a return trip burdensome were unlikely to obtain the ID card and participate in the election.

El Salvador

El Salvador demonstrates one of the other important aspects of diasporas' relationships with their sending countries: hometown associations and their contributions to development. The institutionalization of formal democracy in El Salvador was marked by the 1992 peace agreement between El Salvador's government and the rebel FMLN. The emergence of democracy sparked greater involvement and contributions from Salvadorian emigrants. The main form that this increased interaction took was in the form of hometown associations (HTAs). HTAs are groups in a host country that unite emigrants from the same village, town, or general region of a sending country. They often contribute to development projects and other initiatives in their home region.

One particularly instructive example of a Salvadorian HTA's involvement with its respective hometown that illustrates some of the problematic aspects of HTAs involves the town of Intipucá and the Fundación Unidos por Intipucá. Some observers think that the relationship between these two actors is a model of how diaspora groups can positively contribute to development. Others think that the Fundación's work in Intipucá exemplifies dependence, a lack of productivity, and the exacerbation of social ills.¹⁰⁵ A great deal of money from the Fundación has flowed into Intipucá. However, these resources have been almost entirely spent on projects that benefit the lower socioeconomic classes in El Salvador very little if any at all. The allocation of funds was decided by the Fundación's members at its headquarters in Washington, D.C. Members of the Fundación are relatively well-off, and their interests in Intipucá generally lay in the city center. Accordingly, virtually all of the Fundación's donations and projects have focused on the urban center. This has resulted in a rather glamorous, modern city center surrounded by dismal slums populated by desperately poor Salvadorians. In many ways, HTAs like the Fundación constitute a

⁹⁹ Itzigsohn and Villacres, "Migrant Political Transnationalism," 669.

¹⁰⁰ Ibid., 670.

¹⁰¹ IDEA, 17 and 24.

¹⁰² IDEA, 179.

¹⁰³ IDEA, 187.

¹⁰⁴ Itzigsohn and Villacres, "Migrant Political Transnationalism," 671.

¹⁰⁵ Ibid., 679.

new elite, one that reinforces the existing social structure and contributes money in ways that serve its own interests. In the past, there have not been any legal or institutional measures in El Salvador to ensure equality or the representation of different interests in the construction of development projects funded by the diaspora.¹⁰⁶

Perhaps recognizing the problems associated with “development” projects that only function to serve elite interests, the government of El Salvador has taken steps in recent years to encourage diaspora groups to contribute in a more beneficial, balanced manner. Since 2003, the government has tried to harness Salvadorian HTAs’ interest, engagement, and contributions through the Unidos por la Solidaridad (United for Solidarity) program. Unidos por la Solidaridad is a transnational development program intended to coordinate the actions and resources of HTAs, local communities, and municipal and national governments on economic and social development projects in El Salvador.¹⁰⁷ The key component of the program is a contest in which municipal governments, in conjunction with HTAs, compete for government funds. The contest requires that each competitor develop an investment plan as well as outline a 5-year development agenda.

These measures may be beneficial in encouraging diaspora organizations to truly consider the impact of their intervention in their homelands. If the actions of organizations such as the Fundación exacerbated socioeconomic divisions simply through miscalculations or short-sightedness, then plans such as the Salvadorian government’s hold a great deal of promise. However, if a diaspora organization is actively and knowingly reinforcing elite interests, such measures are likely to be unsuccessful. In this case, the sending country’s government must decide whether this financial inflow is of greater significance than the problems that it creates. If the contributions are deemed to be more harmful than beneficial, the sending government should try to exclude the money. However, this poses a number of problems. The government must distinguish between harmful and beneficial resource inflows, as restricting foreign assistance in general will be detrimental.

Haiti

The case of the Haitian diaspora illustrates a number of dynamics involved in engaging the diaspora, including the role that emigrants can play in undermining the sending state’s government and functioning as a force for political support. When Haiti was under the leadership of President Duvalier, the government largely tried to sever the relationship between Haiti and Haitians abroad.¹⁰⁸ Many members of the diaspora saw themselves as a political alternative to his regime. Although they could have no direct effect on the Haitian political process because they could not vote, the diaspora did not support the Haitian government, voiced its discontent, and sought the ouster of Duvalier.

When Duvalier fell from power, it was generally a result of forces other than the Haitian diaspora’s influence or activities. However, the transition did bode well for the diaspora. The Haitian government began to reverse the anti-emigrant policies of the Duvalier regime and reintegrate the diaspora into the national community. The establishment of the “10th department,” or a 10th administrative district focused on the diaspora was intended to open a channel of communication and engagement between the government and Haitians abroad.¹⁰⁹ Although the government reached out to the diaspora to some extent, it did not extend political rights such as external voting or dual citizenship. Perhaps when democracy and multi-party, competitive elections are fully institutionalized in Haiti the diaspora will be accorded some political rights. The potential political contributions of the Haitian diaspora are evident. When President Aristide was in exile, emigrants provided him with both political and economic support.¹¹⁰ The diaspora can only become more influential as other actors recognize it as potential resource.

India

India is an interesting example of a sending country that, despite having a large and influential diaspora, has tried to exclude its emigrants from its political, social, and economic life. When India gained its independence in 1947, the Indian government’s, and many individual Indians’, stance was generally that those who had left the country

¹⁰⁶ Ibid., 680.

¹⁰⁷ Ibid., 681.

¹⁰⁸ Jose Itzigsohn, “Boundaries of Citizenship,” 1134.

¹⁰⁹ Ibid., 1135.

¹¹⁰ Ibid., 1136.

had forfeit their rights as Indians and should try to integrate into their host countries as much as possible.¹¹¹ In addition, India's closed and staunchly independent model of development during its early years actively discouraged outside influence and involvement, including the potential contributions of the diaspora.¹¹² India was to be governed and recreated by true Indians alone, not those who had abandoned the country and not by other foreign actors.

During the 1960s and the 1970s, India experienced a large labor exodus, as many individuals relocated to find work in Middle Eastern and Gulf countries.¹¹³ However, this emigration was mostly temporary, as the host countries desired short-term labor assistance, not permanent immigration. The brain drain, in which Indian professionals and intellectuals have left the country, has been a separate phenomenon. Both flows could have been manipulated by the Indian government in favorable ways, drawing resources into the country for development, but the negative perception of emigrants proved to be a formidable barrier. Throughout the 1970s and 1980s, India failed to open itself to migrant investment in any significant way, retaining its closed model of development.

At the end of 1990, India's supply of foreign capital was almost entirely depleted.¹¹⁴ This was a huge crisis for the government, and it began to reconsider its stance on foreign contributions. In a desperate move, India developed special concessions for emigrants who invested in the country, but the non-repatriation of capital was still the standard for these investments. By relying on its diaspora, India was able to avoid dependence on the IMF and its conditional loans. Although India granted some benefits to emigrants in this case, its actions to address the economic crisis still represent a missed opportunity to effectively engage the diaspora. Outside investors, including the diaspora, were viewed as a necessary evil and as economic predators. Once the economy recovered beyond the brink of collapse, economic liberalization stalled. The contributions of Indian emigrants, who were

perceived to be corrupted by their host countries, were ignored once again.

Since 1997, economic reforms have resumed, but FDI remains low. This is disappointing, as the Indian diaspora is rather wealthy, successful, and interested in Indian affairs. In recent years, the Indian government has developed a few initiatives that may signal a shift in attitudes toward the diaspora. For example, Person of Indian Origin (PIO) cards have been introduced.¹¹⁵ These cards are the equivalent of a 20-year visa and accord certain privileges to their bearers, including special treatment in buying property and investing in the country. External voting was allowed for the first time in 2004 by post or proxy, but only members of the armed services or civil servants deployed abroad were eligible.¹¹⁶ In addition, the Bharatiya Janata Party (BJP) announced in 2002 that it would begin the process to grant dual citizenship to Indians resident abroad.¹¹⁷ The BJP likely made this move because it is an unabashedly Hindu nationalist organization and recognizes the potential political benefits of doing so, as the Indian diaspora has generally been quite supportive of such organizations. However, the constitutional changes necessary to institute dual citizenship have yet to be set in motion, and it is unlikely that they will be any time in the near future.

Eritrea

The relationship between Eritrea and its diaspora illustrates several important aspects of sending state-emigrant linkages, including the selflessness that emigrants can show in contributing to their homelands and the ways in which previously cozy relationships can sour. Linkages between Eritrea and its diaspora predate the establishment of Eritrea as an independent state. During Eritrea's struggle for independence, the diaspora was mobilized, especially by the Eritrean People's Liberation Front (EPLF) to increase awareness of the conflict, provide political advocacy, and raise money for the war, relief operations, and welfare services.¹¹⁸ These

¹¹¹ Marie Lall, "Mother India's Forgotten Children," in *International Migration and Sending Countries*, ed. Eva Ostergaard-Nielsen (New York: Palgrave Macmillan, 2003), 122.

¹¹² Ibid., 125.

¹¹³ Ibid., 127.

¹¹⁴ Ibid., 128.

¹¹⁵ Ibid., 134.

¹¹⁶ IDEA, 239 and 19.

¹¹⁷ Lall, "Mother India," 121.

¹¹⁸ Khalid Koser, "Long-Distance Nationalism and the Responsible State: The Case of Eritrea," in *International Migration and Sending Countries*, ed. Eva Ostergaard-Nielsen (New York: Palgrave Macmillan, 2003), 174.

contributions were vital in sustaining the independence movement.

Recognizing the importance of the diaspora, the Eritrean government opted for a broad definition of citizenship at independence that was not dependent on residence in Eritrea or an individual's status in his host country. In the 1993 Referendum for Independence, 84,370 votes were cast by Eritrean emigrants. For context, it is estimated that of a total 3 million Eritreans, 1 million are resident outside the state.¹¹⁹ The diaspora was also intimately involved in the creation of the Eritrean constitution. Although emigrants were able to vote in the referendum on independence, this seems to have been a one-shot deal, as they are no longer accorded external voting rights.¹²⁰ As mentioned earlier, shortly after it gained its independence Eritrea instituted a voluntary 2% tax on the diaspora that the majority of emigrants complied with.¹²¹ Eritrea's rapid social, economic, and political reform and concurrent economic development and democratization may bode well for emigrants in obtaining more political rights including external voting.

Eritrea called on its diaspora again during its 1998-2000 war with Ethiopia. The state took three main steps to establish deeper connections with emigrants.¹²² First, it reopened previously shuttered political office in host states to ascertain the demographics of the diaspora. Second, it revitalized the Eritrean Relief Association network. This organization operates under different names in different countries. For example, in the United States it is the Eritrean Development Fund. Third, it initiated an information campaign within the diaspora to spread awareness of Eritrea's needs in the war and the importance of emigrants' contributions.

Financial contributions from the diaspora have been especially important for Eritrea because the state is suspicious of foreign aid. During the war with Ethiopia, Eritrea added significant additional financial requests to the standing voluntary 2% tax. For example, in the United Kingdom Eritrea asked for an additional £1 per day as well as a £500 lump sum for 1999. The government also began issuing

bonds in 1999 and auctioning homes in Asmara, the capital, to the diaspora.¹²³

In recent years, Eritrea has become less enthusiastic about its diaspora. It has become concerned about emigrants' critiques of the government as well as the growing autonomous linkages between the diaspora and individuals in Eritrea. It has also been disappointed in the diaspora's limited success with political advocacy.¹²⁴ Eritrean emigrants have also developed negative perceptions of the state. Many emigrants are experiencing a growing disillusionment and a feeling of exploitation.¹²⁵ The excessive financial demands of the government have strained the relationship between emigrants and the state. There is overwhelming social pressure within diaspora communities to contribute generously to the state, with individuals often checking on each other to ensure that everyone is giving. This pressure, combined with emigrants' desire to contribute to their family members in Eritrea as well, has simply been too much for the diaspora to handle in many cases.

States face a number of limitations in efforts to increase their engagement with their diasporas. First, the ever broader dispersal of emigrants can strain resource-strapped states' abilities to effectively reach these populations.¹²⁶ Second, the principle of territorial sovereignty restricts states' power in dealing with their diasporas.¹²⁷ Although they may communicate and develop relationships with their emigrants, they generally do not wield any coercive power over these individuals. This can be especially problematic when the relationship between the sending state and its emigrants is antagonistic rather than synergistic. Although there are a number of issues associated with engaging and extracting resources from diaspora populations, the importance of doing so will likely grow in the future as migration flows continue to increase. Sending states must assess the resources they can potentially extract from their diasporas and plan their engagement strategies accordingly.

¹¹⁹ Ibid., 172.

¹²⁰ ACE Project, "Comparative Review of Voting from Abroad," <http://aceproject.org/ace-en/topics/va/comparative-review>

¹²¹ Koser, "Long-Distance Nationalism," 175.

¹²² Ibid.

¹²³ Ibid., 176-7.

¹²⁴ Ibid., 178-9.

¹²⁵ Ibid., 180.

¹²⁶ Barry, "Home and Away," 6.

¹²⁷ Ibid., 6.

Bibliography

- ACE Project. "Comparative Review of Voting from Abroad." Accessed April 8, 2011
- Adamson, Fiona and Madeleine Demetriou. "Remapping the Boundaries of 'State' and 'National Identity': Incorporating Diasporas into IR Theorizing." *European Journal of International Relations* 13 (2007): 489-526.
- African Diaspora Policy Centre. "Building Institutional Cooperation Between the Diaspora and Homeland Governments in Africa: The Cases of Ghana, Nigeria, Germany, USA, and the UK." *ADPC Policy Brief* (June 2010). Accessed April 8, 2011
- Barry, Kim. "Home and Away: The Construction of Citizenship in an Emigration Context." *New York University Law Review* 81 no. 11 (2006). Accessed April 8, 2011
- Brinkerhoff, Jennifer M. "Exploring the Role of Diasporas in Rebuilding Governance in Post-Conflict Societies." In *Africa's Finances: The Contribution of Remittances*, 241-264. New Castle upon Tyne, UK: Cambridge Scholars Publishing, 2008. Accessed April 8, 2011
- Collier, Paul. "Economic Causes of Civil Conflict and Their Implications for Policy." In *Leashing the Dogs of War: Conflict Management in a Divided World*, edited by Chester A. Crocker et. al.: 197-218. Washington D.C.: United States Institute of Peace Press, 2007.
- Collyer, Michael and Zana Vathi. "Patterns of Extra-territorial Voting." *Sussex Centre for Migration Research Working Paper* T22 (October 2007). Accessed April 8, 2011.
- Gamlen, Alan. "Diaspora Engagement Policies: What are They, and What Kinds of States Use Them?" *Centre on Migration, Policy, and Society Working Paper* no. 32 (2006). Accessed April 8, 2011
- de Haas, Hein. "International Migration and National Development: Viewpoints and Policy Initiatives in Countries of Origin: The Case of Nigeria." *International Migration Institute: University of Oxford* (June 2006). Accessed April 8, 2011
- International Institute for Democracy and Electoral Assistance. "Voting from Abroad: The International IDEA Handbook." *International Institute for Democracy and Electoral Assistance* (November 2009). Accessed April 8, 2011
- Itzigsohn, Jose. "Immigration and the Boundaries of Citizenship: The Institutions of Immigrants' Political Transnationalism." *International Migration Review* 34 no. 4 (2000): 1126-1154.
- Itzigsohn, Jose and Daniela Villacres, "Migrant Political Transnationalism and the Practice of Democracy: Dominican External Voting Rights and Salvadorean HTAs." *Ethnic and Racial Studies* 31 no. 4 (2008): 664-686.
- Keely, Charles B. and Bao Nga Tran. "Remittances from Labor Migration: Evaluations, Performance, and Implications." *International Migration Review* 23 no. 3 (Autumn 1989): 500-525.
- Kinwall, Catarina and Ted Svensson. "Hindu Nationalism, Diaspora Politics, and Nation-Building in India." *Australian Journal of International Affairs* 64 no. 3 (2010): 274-292.
- Lall, Marie. "Mother India's Forgotten Children." In *International Migration and Sending Countries*, edited by Eva Ostergaard-Nielsen, 121-139. New York: Palgrave Macmillan, 2003.
- Leblang, David. "Harnessing the Diaspora: The Political Economy of Dual Citizenship." *University of Virginia* (February 2010) Accessed April 8, 2011.
- Lyons, Terrence and Peter Mandaville. "Global Migration and Transnational Politics: A Conceptual Framework." *George Mason University Project on Global Migration and Transnational Politics* Working Paper no. 1 (2008). Accessed April 8, 2011.

Martinez-Saldana, Jesus. "Los Olvidados Become Heroes: The Evolution of Mexico's Policies Towards Citizens Abroad." In *International Migration and Sending Countries: Perceptions, Policies, and Transnational Relations*, edited by Eva Ostergaard-Nielsen, 33-56. New York: Palgrave Macmillan, 2003.

Newland, Kathleen. "Voice After Exit: Diaspora Advocacy." *Migration Policy Institute* (2010) Accessed April 8, 2011.

Organization for Economic Cooperation and Development. "International Migration Remittances and Their Role in Development." *International Migration Outlook* (2006) Accessed April 8, 2011.

Ostergaard-Nielsen, Eva. "International Migration and Sending Countries: Key Issues and ." In *International Migration and Sending Countries*, edited by Eva Ostergaard-Nielsen, 3-30. New York: Palgrave Macmillan, 2003.

Portes, Alejandro, Luis E. Guarnizo, and Patricia Landolt. "The Study of Transnationalism: Pitfalls and Promise of an Emergent Research Field." *Ethnic and Racial Studies* 22 no. 2 (1999): 217-237.

Sachs, Susan. "New York, Citing Security, Rejects Mexican ID Cards." *New York Times*, December 28, 2002. Accessed April 8, 2011.

The World Bank. *Migration and Remittances Factbook* 2011. Washington, D.C.: The World Bank, 2011. Accessed April 8, 2011.

The Politics of Environmentalism in Modern Day China

William Labate

Introduction

Over the past few decades the People's Republic of China (PRC) has experienced breathtaking growth and development. Millions of Chinese have risen out of poverty while China has experienced one of the largest rural to urban migrations in history. Most economists predict that China's growth will outpace even the dynamic United States economy. There is, however, a tradeoff for all of this growth, urbanization, and modernization within China. A byproduct of this prosperity has been the destruction of China's environment for the sake of continued economic growth. The most obvious issue is that China is the largest emitter of greenhouse gasses in the world, which of course contributes to global warming. While greenhouse gas emissions are not a contentious issue for the average Chinese citizen, there are many issues which are, ranging from large scale air pollution to local land confiscations. The focus of this paper are these underlying environmental issues which have large implications for individuals and civil society. Because these issues are so prevalent, individuals and groups seek to remedy them through environmental activism. In any democratic country this activism would take the shape of active NGOs and individuals lobbying their government to implement effective environmental policies. By the nature of China's political system, this activism takes quite a different shape.

The objective of this paper is to examine the limits of environmental activism set by the government and how Chinese civil society attempts to operate within those guidelines. The paper wants to show that the PRC does not operate with a rigid, top down structure, and that there are in fact many variations in priorities, actions and outcomes between the different levels of the Chinese government. It examines the barriers that activists encounter and the most successful methods that they employ to overcome them. Using these methods this paper suggests several ways that activists might want to employ in order to be successful in the future. Since China is not an ideal place for environmental activists, this paper also attempts to offer several policy prescriptions for the PRC to better address environmental and other

concerns both before and after they emerge. Some of these prescriptions are unlikely to be implemented at the present time, but can be pursued in the long run while still maintaining China's current style of government.

In China, environmental activism must be carefully constructed to fit into the Chinese government structure to avoid being rebuked. The Chinese government is dominated, of course, by the Chinese Communist Party (CCP) which operates party cells at every single level of government. The central communist party bureau sets economic, political, and social guidelines for the rest of the country to follow. To a limited extent, the central party sets environmental policy as well for issues of national concern like carbon intensity, certain standards, reforestation, etc. Still, many environmental problems are of a regional and local nature and are solved locally which we will see with our case studies (with the last being an exception). What this paper also looks at is how the PRC is not a monolithic one-party state with directives coming from the top and reaching the bottom in the same form. Every regional and even local party branch has different styles and priorities as will be seen in this paper. A directive from a higher level of government doesn't always get implemented the way it is intended. The lower level of government might want to tailor it to further their own priorities or ignore it all together if it contradicts their goals. This kind of internal dissent cannot be tolerated for issues of high priority, but for low priority issues it is permitted.

Environmental issues are of a relatively low priority compared to issues like economics or security. When George Bush asked Hu Jintao what kept him awake at night and he replied, the worry of "creating 25 million new jobs a year."¹ If it means ignoring the protection of the environment in order to prevent the creation of millions of jobless youth, the Chinese government will do so. Yet, meeting these economic targets have been, and will be the highest priority in China for a while. Creating these jobs will be worth the environmental externalities because the CCP judges unemployment as the most dangerous issue. Despite this, there has been a general backlash from

¹ "Reassessing George Bush."

the Chinese citizenry for not having stricter environmental standards. The national and regional governments have attempted to placate these demands by implementing stricter standards and improving environmental quality but only to a limited extent. If China's environment degrades in a more rapid fashion, citizen displeasure might move environmental issues to a higher priority.

Due to the nature of its political system, the PRC and its citizens must reconcile how environmental activists can operate in an authoritarian, one-party state. The environmental regulations that the CCP write are the law of the state and are expected to be followed. By challenging those laws or lack thereof, environmental activists might be seen as opposing the CCP and by extension, the state. Even by framing activism as non-political concerning only environmental issues, it is easy to see how that can be extended to being considered a political issue as well. Despite this, the PRC has loosened its grip on civil society since the days of Mao. In the past few decades, civil society and the press have been given much more freedom to operate in many more areas than before. Though there are still many restrictions that the government places on communication and the internet that restrict NGOs from organizing. Additionally, in most places NGOs have to be pre-approved in order to operate which severely hampers the emergence of new NGOs around various issue areas. If an issue is too controversial or the government thinks that the organization will pose a threat to its rule, the government does not have to grant it recognition. This lack of legal status gives the government an excuse to suppress their activities if needed. It is still a challenge for NGOs and individuals to pinpoint the boundaries that the state sets forth for them in their activism. Where the state puts these boundaries and the methods for how civil society determines their location are what the case studies in this paper examine in great detail. I will explore these issues in the following three case studies of environmental activism in China.

The first case study examines the Green Garden conflict in Shanghai during the 1990s. In two instances, the local government wanted to build a structure in the middle of the green area in the Green Garden Village. According to the municipal regulations, green area is supposed to be preserved for the people of the village and not to be built up. The residents of the village attempted to assert their environmental rights by applying pressure to the local and city governments as well as different

regulatory agencies to revoke the permit. The local government, however, sought to prevent the protesters from succeeding by manipulating existing regulations. This case study reveals how a local government interacts with an organized and coordinated environmental protest movement who uses their own laws against them. It also is intended to reveal how far protesters can go before they go outside the limits of non-political issues. In addition to the government-protester interaction, there are also several rule of law failures on the part of the government and bureaucracies which are examined. The use of journalists and print media also play a role in channeling support for the movement. Finally several policy prescriptions are issued to encourage consistent applications of regulations and clearer land laws.

The second case study moves to the municipal level of government of Shanghai and the conflict over the Jiangwan Marshlands. The government wanted to drain the marsh and develop the area in order to accommodate the fast growing city. There were many Shanghai residents who did not want the marshlands to disappear at all since they are a vital green space which offers a prime place to bird watch. The fight to preserve the marshlands was led by an illegal and unregistered NGO, the Shanghai branch of WWF. The residents, many of them who were part of this NGO organized meetings and petitions against the building. They primarily used the internet as their chief medium of communication to organize but due to their illegal status, the government was able to step in and shut them down. Ultimately their efforts failed with only a small part of the marshland preserved and the rest build on. This case study is intended to show how the government handles dissent from illegal NGOs at a larger level of government. It also highlights how the internet can be used to organize protesters and how the government responds to that medium. Finally, it shows how the government out rightly prioritizes economic growth over the preservation of green space.

The third study is the dam project on the Nu River in which a legal NGO, the Green Environmental Volunteers (GEV), managed to prevent this project from being constructed. The study examines the favorable NGO climate in Beijing and how NGOs can operate with a reduced fear of persecution. Establishing bonds and connections were crucial in the study since the GEV was able to collaborate with NGOs from the other side of the country to collect data and raise public opinion. Another interesting factor was the

fact that the Ministry of Environmental Protection was on the side of the NGOs and assisted them in the form of giving information. The fissures between different parts of the central government will be analyzed as well as effective strategies to increase cooperation between NGOs.

Several factors went into the methodology of selecting these three case studies. At the most general, the cases were selected for their different levels of analysis with the Green Garden village being at the most local level, Jiangwan marshlands at the municipal/regional level and the Nu River at the regional/national level. It is necessary to see how at all levels the government interacts with environmental activism and how different priority levels affect that interaction (i.e. issues of national priority are more critical than regional/local). The form of advocacy is also another critical factor into determining the case studies. The Green Garden case was driven primarily by a few motivated individuals, the Jiangwan case with a loosely organized illegal NGO, and the Nu River with recognized and more organized NGOs. The rising importance of the internet is also important to examine with it playing a big role in the Jiangwan study. Each study examines a difficult hurdle that Chinese civil society has to overcome like rule of law failures (Green Garden), NGO recognition and internet censorship (Jiangwan), and NGO collaboration (Nu River). Two were big successes but another was a failure (Jiangwan) which is significant because these protests do fail on many occasions. What they all have in common is the PRC putting economic issues ahead of environmental ones. They all examine fissures and divergent interests at the different levels showing that the PRC is not a unitary actor both within and between levels of government. Each case study shows how the government attempts to clamp down on activists and stifle their dissent with varying results. Finally each one identifies the pitfalls that activists have to avoid in order for them to be effective in an authoritarian state.

Case Study #1: Individual Activism in the Shanghai Green Gardens

This first case study examines the conflict between residents and the local government over construction in the green space of the Green Garden New Village in Shanghai. In an effort to address environmental decay and degradation inside Chinese cities, several municipalities embarked on urban renewal campaigns during the 1990s. The

Green Garden New Village in Shanghai was one such initiative designed to resettle residents from the city center to new residential communities closer to the city's periphery. Regulations established when constructing these villages mandated that 10.16 percent of the village's land area is comprised of green space. Green Garden village was comprised of 12 apartment blocks with a total of 10,000 residents surrounding a 8,000 square meter park. Residents were relocated from their dwellings in their traditional neighborhoods, or "*lilong*," and moved to these new villages. Many of the 12 housing blocks consisted of residents from several different *lilong* but a several of the blocks had been "collectively resettled" with residents all from the same *lilong*.²

The individual villages of the Shanghai municipality are composed of a series of different divisions managing affairs at the most local level. Theoretically, the most powerful citizen organization is the Resident's Committee comprised of local residents which is supposed to represent their interests to the local government. In reality, the members of this organization also must be members of the local cells of the CCP so they largely represent the party's interest instead.³ The Residents' Committee, which is supposed to be politically autonomous, is in actuality held accountable to the Street Council which is the most basic form of local governance. This is a good example of how the CCP uses organizations which are supposed to be representing citizen interests, to represent their interests. With housing being commercialized and many of the tasks of a landlord moving from the responsibility of the government to real-estate management companies, the city of Shanghai decided to set up home-owners' committees in 1997. These committees were assigned one to each block and were designated to handle concerns between residents and the real-estate management company. The real estate management companies, although seemingly independent from government, exist to serve the government's interests and are extensions of their power. It is a good example of China's "Red Capitalism," or state direction of economic entities. Beneath all of these organizations are informal community associations of citizens who live in the

² Zhu, Jiangang. "Not against the state, just protecting residents' interests," 155.

³ Ibid.

area and have common interests, like the Elderly Association.⁴

The first major occurrence of green space destruction in Green Garden came when a real-estate developer obtained a permit to build a structure in the middle of the village park in 1993. After the construction began, it was discovered by a group of elderly woman who spread the news through their informal associations within the community. They reported this violation to the Residents' Committee who then notified the Street Council and the real-estate developer. The residents, through their local associations, wrote a series of complaint letters to the company which were all ignored. At the time, the home owners' committees had not been created so the residents turned to another village resident named Shen Xin who agreed to help them.

Shen carefully constructed an advocacy strategy that was both effective in getting the residents' argument across and not getting him arrested. He needed to carefully define a reason to advocate for the protection of the green space that did not appear to go against the party or the government so their actions would not be suppressed. Shen framed his complaints in order to convey that residents were seeking to "protect their environmental rights" according to the laws passed by the government.⁵ To back up these rights, Shen used an "authenticating strategy," which meant that he was not disagreeing with the official policy on environmental issues; he was simply compelling the government to adhere to their own laws. For example, the Shanghai Green Administrative Regulations explicitly say that "Building in public green space is forbidden."⁶ The issue in this case was not that there was a lack in regulation protecting the environment; it was that existing regulations were not being enforced. Shen would use these regulations when discussing issues with officials and committees making it difficult for them to contradict government laws. With the lack of response from the Street Council, Shen took his complaints to the local Bureau of Urban Planning and the Bureau of Park Management who were charged with protecting the environment. Unfortunately for Shen, these two bureaus did not want to confront the other branches of the local government even though they supported his cause. This is a perfect example of rule of law failure

since the bureaucracy should be acting according to the law and not trying to interpret the wishes of the party. Since local remedies were not producing the desired outcome, Shen managed to convince the residents to file a "collective complaint" to the Shanghai municipal government.⁷ What he was trying to do was draw attention to these abuses by the local government and show that enough citizens were displeased with this failure.

Shen decided to become bolder with his struggle to preserve the park as the battle went into its second year. Restrictions on the media had been loosened during the past few decades on issues which are not too politically sensitive in nature. Shanghai's newspapers began publishing stories on the Green Garden conflict which eventually came to the attention of many high municipal officials. The vice-mayor issued a statement that the situation should be handled "according to the law."⁸ Though this was not a direct order to end the project, the Bureau of Park Management interpreted it as enough of a justification to order a freeze of further construction. Despite the freeze, the developer continued to build on the site in secret but the residents quickly discovered this violation. At this point, Shen made his most risky gamble yet when he ordered a few of the activists to break into the construction site and take down all of the scaffolding. When asked why he did such a thing, Shen replied that "The government had already forbidden the construction but they still went on, so what they were doing was illegal. We rushed into the site just to stop the illegal work. That's why we dared to do it."⁹ By framing it in this way, Shen was able to portray the struggle as the people against the development company and not the people against the government (even though they were essentially the same entity). After this incident as well as subsequent stories in the press, the Bureau of Urban planning revoked the construction permit and forced them to leave the park.

Even with their initial victory to preserve the park, the residents of Green Garden had to fight for their environmental rights once more in 1997, but this time it was against the local government. The original plans for the park included the opportunity for the construction of a small, 135 square meter community center for the elderly. Despite these clear restrictions, the district government and the Street Council wanted to build a 2,600 square meter

⁴ Zhu, 156.

⁵ Zhu, 157.

⁶ Ibid.

⁷ Ibid.

⁸ Zhu, 158.

⁹ Ibid.

center for “old cadres” who were part of the movement which helped found the PRC.¹⁰ Upset at this seizure of park lands, the residents protested that the park belonged to the community and were protected by the state under the law. The district government claimed that the park belonged to the state and they had the right to do with it what they wished. After the homeowners’ committee failed to reach an agreement with the Street Council, they once again turned to the media for support. After hearing about the story in the news, the municipal government put pressure on the district to compromise with the residents. Both the Street Council and the Homeowners’ Committee agreed to the construction of a larger community center in the park than what was detailed in the regulations, but only at the size of 650 square meters and not for cadres. The Street Council however, did not honor the agreement and decided to build a center for the cadres anyway at a size of 1,960 square meters.

After discovering that their agreement was going to be disregarded, the residents were outraged. Some wanted to take radical action and destroy the construction site all together. This might be seen as a direct attack on the local government and a precursor to further violent actions. Shen knew this would turn many of his allies and the higher levels of government against the movement so he took a different approach. He led about 100 residents to the construction site to negotiate with a government representative. When nobody came to talk, Shen ordered the residents to pull down the barrier wall around the site and do nothing else. In a very precise and orderly manner Shen hoped that the local government would be “afraid of the resident’s potential ‘violence’” but avoided any outward example of real violence.¹¹ Seeing this display, the Street Council temporarily stopped construction only to continue soon after.

As the builders continued construction on the expanded site, a few of the younger residents began to circulate a petition to end it. The petition itself wasn’t risky however, the residents wanted to submit it to the municipal authorities around June 4th, a sensitive date because of the June 4th movement in China during 1989 democracy protests. Shen thought that this was too politically risky and wanted to get the government’s attention but “not let them lose face.”¹² They submitted their petition to the government office in the evening in

an attempt to avoid public attention. In this way, the Shen and his followers wanted to not seem like they were attempting to blatantly muster public support in challenging the government. The Street Council struck back and declared that a petition around the period of June 4th was illegal. They organized a public meeting with the Residents’ Council and retired members of the Communist party where they declared that Shen and his followers were acting out of selfish interests and not of that of the community. The Street Council even called the police to come and surround the building which served to intimidate the residents. Despite Shen’s efforts to “depoliticize the matter, the Street Council consciously treated the conflict as a political question.”¹³

This intimidation by the police created fissures between the 12 homeowners associations and threatened the entire movement. Those who lived in the same *lilong* before they were moved showed the most commitment to the cause demonstrating how far old community ties went.¹⁴ Shen wanted to press on with the strategy of depolarization even after the Street Council tried to bribe him. He decided to continue to go through legal route by making repeated complaints to the Bureau of Park Management and the Bureau of Urban Planning about the illegal construction even though they were still under pressure from the street council not to act. He once again reached out to the media where he portrayed the issue as an environmental conflict and not as one against the government. Through this channel, the wider public as well as many municipal officials were made aware of the Green Garden issue. Shen even went further to bring a lawsuit against the development company in order to compel them to follow the law. In this way he hoped “to establish the rule of law instead of the rule by administrative government.”¹⁵ By focusing on the application of the law, Shen sought to use that as his cover from being accused of challenging the state. Shen claimed that the residents “defend the law. We are the people, so we stand on the side of the state. We know we are not against the state, but are just protecting our interests.”¹⁶ After facing all of this opposition and pressure from above, the local government decided to abandon the project.

¹⁰ Zhu, 159.

¹¹ Zhu, 151.

¹² Zhu, 161.

¹³ Ibid.

¹⁴ Ibid.

¹⁵ Zhu, 163.

¹⁶ Ibid.

Analysis and Policy Prescriptions

The Green Garden incidents highlight several trends in how the government and citizens interact over environmental discourse in China. First, the citizens have a much larger amount of leeway to protest environmental issues which are usually of low political significance. In this particular case, the construction of a senior center in the middle of a village park does not have political significance outside the local level. Shen's more radical actions like taking down the scaffolding and tearing down the wall around the construction site would not be tolerated if the issue in question was more sensitive. It did help that the city government did give tacit support for the goals of the protesters. Though Shen's goal was to depoliticize the environmental issue in question, the July 4th petition threatened to contradict that stance. Shen was still able to avoid any uncomfortable statements near the anniversary but the government still tried to politicize the matter. This politicization of non-political environmental issues was an attempt by the street council to drag the issue into the forbidden zone of political dissent. After failing to effectively contradict Shen since his legal argument was stacked against them, they attempted to halt his protests from this approach. Shen was able to drag the protest out of that zone and back into that of non-political dissent. It was clear to Shen that in China, discourse is only tolerated within those limits of non-political dissent. Once it strays outside those boundaries it can be interpreted by the government at higher levels as a threat or challenge to their rule. Shen accurately knew that by talking about these issues in a non-political light was the only way to do so without fear of retribution. His success at bringing the issues back into this light was the reason why it ultimately succeeded and was not shut down by the police.

The second trend was that the use of official channels and the media to effectively register dissent. Shen consistently quoted government regulations and the citizen's environmental rights as justification to halt the destruction of the green space. By using the PRC regulations while talking to officials, Shen placed them in a difficult position since they would have to contradict laws written by the party and government. This authenticating strategy is useful since sometimes the laws for proper environmental protection are in place but they are simply not enforced properly. Shen also used completely legitimate petitions and collective complaints to alert the higher level of government

that these abuses were occurring which led to pressure on the street council from above. Petitioning against abuses of the law by the government is actually encouraged by many leaders in China. Just recently Prime Minister Wen Jiabao appeared at the nation's top petition bureau to, as the New York Times describes it, compel "citizens to criticize the government and press their cases for justice" when laws are being violated.¹⁷ It should be noted that the petition bureau is not intended to make the government more liberal but instead make it more efficient. This criticism would have never been tolerated in the Maoist days of China where any criticism of government would be seen as a challenge to the state. Though today, it is still doubtful that issues of real political or economic importance would be accepted by the government. It is easy to sweep away Wen's statements as political posing and rhetoric but for environmental issues the petition bureau can assist residents to a limited extent. Alone, the petitions that Shen submitted against the building weren't very effective but combined with media assistance they managed to accomplish his goals.

In both of the Green Garden conflicts the press played a critical role in disseminating information and for turning the people and municipal government to the side of the protesters. Though authoritarian, the PRC gives its press a certain (albeit short) amount of leeway to operate and report on stories. The media performs a certain amount of "self-censorship" to avoid potential "political minefields" while still being allowed to report on somewhat sensitive topics.¹⁸ Most politically sensitive issues are still off limits but since environmental issues are of a relatively low political importance, in both cases the media decided to pick up the story. In both Green Gardens cases a law, written by the CCP, was clearly being violated and the residents had legitimate grievances. Even in the case where the local government was involved, these newspapers circulated city-wide. By going against the paper, the local government would attract unwanted attention for its law-breaking from the public and from the municipal government. While the media cannot go against the party line it is usually acceptable to expose government abuses at the local level. This media publication was able to

¹⁷ Barboza, David. "China Leader Encourages Criticism of Government."

¹⁸ Tong, Jingrong. "Press self-censorship in China: A case study in the transformation of discourse," 593.

attract enough attention from the municipal authorities that they in fact did apply pressure to the local government to compromise with the protesters. At the end of the first case after the Vice Mayor saw the news stories about the issues he even issued a veiled statement to support the protesters argument. The media also exposed the Bureau of Urban planning which after much reporting withdrew its permit for the building showing that the media has influence on the bureaucracy as well. After Hu Jintao rose to the post of CCP General Secretary in 2002, he wanted to promote more transparency in government through the media. It was Hu's desire for transparency that actually encouraged media to "investigate allegations of corruption and mismanagement of local affairs by local officials."¹⁹ Since these reforms took place after the Green Garden protest, they can be seen as the continuation of the expansion in the range that the media can operate in China. By no means should this be viewed as liberalization since the arena of political reforms is still out of the questions. The goal of these reforms is to make the government more efficient and not necessarily to give the press more freedoms. Though this means that now it is even easier to report on environmental violations in local government, the media is more constrained when criticizing existing regulations or the lack thereof. The media is always under the rather ambiguous Article 51 of the Constitution of the People's Republic of China which states that "The exercise by citizens of the People's Republic of China of their freedoms and rights may not infringe upon the interests of the state."²⁰ This is another vaguely worded clause that the government can cite when they want to limit the publication ability of journalists. In order to improve governance overall and encourage truly effective investigative journalism the CCP needs to loosen its grip on the press. If a particular project or program of vital national interest has problems with corruption or harms the environment, the media would be less likely to cover it because of fear of retribution. The media also might misjudge the "minefield" and may not even cover some stories where their stories might be totally in the bounds to report. Unless changed, the way that journalism laws are in place currently discourages the media from publishing

stories which might lead to better governance and environmental protection in general.²¹

The third trend that was prevalent was the use by the local government of the ambiguous wording of laws and ignoring regulations to try to get their way. Ignoring regulations works in many instances around China when the opposition is not as organized or as cunning as Shen's movement. This rule of law problem can be abused in order to ignore environmental issues for the sake of progress and economic growth. China's land laws also open up more maneuverability since their vague definition of land ownership allows for wide interpretation. It is really up to the citizens to call the government out on their abuses in order for the problem to get fixed in many cases. In other instances, the project may be of such national or regional importance that no amount of protest will convince the government otherwise since it can many times fall back on legal ambiguity if no law specifically forbids their actions.

Both incidents in the Green Garden case reveal several weaknesses of the rule of law on the local level in China which must be addressed in order to avoid future protests. Pan Wei identifies 5 areas on which he bases rule of law; a neutral civil service, an autonomous judiciary, social consultative institutions, and freedom of speech.²² The most glaring failing of the rule of law in this first scenario was the lack of a neutral civil service. The local Bureau of Urban Planning and the Bureau of Park Management were constantly trying to interpret the will of the party rather than apply the regulations that were clearly in place. In general, this leads to arbitrary application of the law by the local bureaucracies both to the advantage and disadvantage of the municipal government. On one side, the government is disadvantaged because the regulations to preserve green space are being ignored. The whole purpose of the Green Garden village was to combat urban decay and lead to a green renewal of Shanghai by promoting more green space. When the local Street Council ignores regulations, it harms the goals of the municipal government and undermines their objectives. The advantage to this weakness in rule of law for the local (and municipal) government is that it can ignore the regulations when it is in their interest to do so. If an important structure needed to be built,

¹⁹ Rawnsley, Gary D.. "Media, Internet and Governance," 124.

²⁰ People's Republic of China. *Constitution of the People's Republic of China*.

²¹ Rawnsley p.125

²² Peerenboom, Randall. " A Government of Laws: democracy, rule of law and administrative law reform in the PRC," 47.

the party officials in the local government can simply put pressure on the civil service bureaucracy to issue them a building permit. Not only does this undermine the goals of the regulations, but it also leads to dissatisfaction among residents, many of which might not be as well organized as Shen to fight this. This dissatisfaction was prevalent during the Cultural Revolution due to the arbitrariness of regulations which resulted in many post-Mao legal reforms to encourage consistent application of the law.²³ These reforms have gone far to ensure much consistency at the national and regional levels, but not at the local levels where they are harder to enforce due to a lack of oversight.²⁴

What the CCP should focus on is separating the party from government in order to ensure the neutrality of China's civil service. Since Wei's 5th criteria for effective rule of law, freedom of speech and to a larger extent, democratic accountability, is currently out of the question, neutrality remains a realistic near-term possibility. This isn't such a radical proposal since "the party has announced as far back as 1978 11th party congress that they intended to separate the party from government."²⁵ The party would still be able to maintain ultimate control since they would still write all of the regulations in the first place. This would mean that departments like the Urban Planning Bureau would enforce regulations to the letter no matter who is involved. Both bureaus involved in the first Green Garden case had to try to interpret the wishes of the vice-mayor when he said the situation should be handled "according to the law." Under a neutral civil service, there would be less second guessing and tiptoeing around officials for fear of retribution. If the Chinese government wants more consistent adherence to regulations across the board they need to reign in these local authorities who repeatedly flout the rules. Not only would this consistency promote more economic growth and foreign investment, it would also go further in meeting the demands of Chinese citizens. Many Chinese do not say that they are dissatisfied with their form of government rather they "increasingly expect the legal system to protect their rights" according to Chinese law.²⁶ The party would still be able to set the rules of the game by making this law but as long as the agencies and citizens play within those rules there should be fewer issues. Despite the

benefits, neutrality has faced many obstacles since it is hard to enforce at the local levels and the CCP enjoys its ability to selectively apply policy when it suits them.

One of the several reasons that issues like those in the Green Gardens case arise, is because of the ambiguity of China's land laws. The government has many reasons to impose a "deliberate institutional ambiguity"²⁷ when crafting the laws. If this ambiguity is combined with the inconsistent enforcement of the laws which clearly delineate land ownership and land rights, the government has more room to adapt their policy to meet their own needs. The phrase "collective ownership" is deliberately left out of many places in the land laws in order to avoid public unrest when the land is used for the purposes of the state.²⁸ This concept of state ownership of land is a departure from the Mao-era policy of registering land under the local production team.²⁹ Now many of the land use laws state that the land belongs to "the people" which the government interprets as "the state" in cases like the Green Garden. This has allowed local governments to take ownership of what should be resident owned land and use it for various development projects. In the case of Green Garden, residents exposed what the local government was attempting to do and used it as a rallying point to turn the public against it.

Luckily for the residents, the government backed down at the end of the second case and the construction was halted. This clearly might not happen in other cases even when residents use similar strategy to Shen's. What happens when the local government is less sympathetic or the central government is involved? As we saw with Green Garden, there was a more radical wing that had a desire for more extreme action which Shen had to temper in order to achieve his goals. That radical side can easily be the one calling the shots in an alternate scenario. The intentional ambiguity of China's land laws can be "potentially explosive source for future social conflict."³⁰ It is up to the Chinese government to make sure that the rule of law is more effectively enforced at the local levels. They also need to clarify their land laws so that ownership is more clearly defined and these

²³ Peerenboom, 48.

²⁴ Ibid.

²⁵ Ibid.

²⁶ Ibid.

²⁷ Ho, Peter. "Who Owns China's Land? Policies, Property Rights and Deliberate Institutional Ambiguity," 400.

²⁸ Ho, 401.

²⁹ Ibid.

³⁰ Ibid.

conflicts do not erupt over preserving green space which should be publically owned. We have seen the bounds in this scenario for how far the government let the protesters proceed. The protesters operation within the laws and used the laws as their justification for their actions. Even in a society where the rule of law is weak it is hard for the government to ignore that these laws were clearly being broken. When the local government failed to stop the protesters they tried to turn the issue into a political matter. By trying to drag the issue of what is allowed into the area of what is not allowed, the government wanted to delegitimize the protests but failed to cast the protests in that light. By the end of this case study it is clear how the government sets the political arena and how far the people can go without retribution.

Case Study #2: Illegal NGO Activism in the Shanghai Jiangwan Marshlands

This second case study expands its scope and examines the Jiangwan marshlands issue at the regional level during 2002. Shanghai one of the largest cities in China and the world with a population of over 17.8 million people.³¹ Even more so when compared to the rest of China, Shanghai has gone through tremendous economic growth transforming into an economic hub for the rest of the country. Many politicians use municipal office in Shanghai as a stepping stone for higher positions within the central government based on their performance. Criteria for these promotions include wealth creation and economic growth so these goals quickly became the top priority for policymakers. Though impressive, this economic growth leads to many negative externalities with environmental destruction being one of the most prominent. This has led to the creation of many advocacy organizations throughout the city designed to address environmental and other concerns. In 1999 the very well-financed Shanghai Bureau of Civil Organizations was created as an extension of the Shanghai Bureau of Civil Affairs. This Bureau is charged with identifying and suppressing (if need be) those organizations and individuals who might challenge the aims of the government. If a civil society organization wants to be recognized, then they have to apply to the bureau for approval before it can exist legally. The government uses this ability as a tool since they can “prevent the development of NGOs by not

providing them legal status.”³² With all of these barriers and regulations it is very difficult to maintain and operate NGOs in Shanghai.

Though in this case study Shanghai is portrayed as being environmentally irresponsible regarding the marshlands, the municipality has attempted to improve the environment to a limited degree. The municipality made great strides in improving water quality, air quality, waste disposal and greenery since the 1990s and hopes to become an ecological city by 2020.³³ Enforcement of environmental regulations is left to the Shanghai Environmental Protection Bureau (SEPB). The SEPB has the ability to compel business to follow regulations or be fined like it has done to thousands of entities since its creation. It uses some of this revenue to invest in environmental protection inside the city. In addition to this bureau the municipality has set up an “Environmental Protection and Ecological Construction Coordinating Committee,” the first of its kind in all of China.³⁴ City officials of the highest levels sit on the committee including the Mayor and the Vice-Mayors. This makes implementation of the committee’s decisions fairly efficient since government agencies know that these decisions come directly from the top. Shanghai has been ranked number one among all other cities in China for its environmental protection efforts. All of these things, though still falling short of what is needed in order to guarantee a safe and clean environment for Shanghai’s residents, are still significant since they go further than what is implemented in other areas of China. The municipal government has made a concentrated effort to improve the environmental status of the city. The money used for environmental protection can go to other places that would promote development showing that Chinese political interests are more complex. The fact that the highest officials in the city have put the issue as a fairly high priority is also significant since in this case study it seems like the opposite is being done. It provides a basis for advocates to negotiate with the city for tougher environmental protection rules since there is a clear precedent for them. Though as seen with this case study, there are limits to how far Shanghai goes to protect the environment.

Despite the restrictions, many NGOs as well as many Environmental NGOs (ENGOS) exist within Shanghai. Most of the ENGOS that do exist

³² Xie, 129.

³³ Xie, 131.

³⁴ Xie, 132.

³¹ Xie, Lei. “Environmental Activism in China,” 128.

are permitted to do so because they do not pose a threat to the government. The ENGO that we are specifically studying in this case is the World Wildlife Fund (WWF) Shanghai. Globally, WWF has many networks of supporters in many different countries and regional branches. Its main goals are the restoration and, as it is in this case, conservation of wildlife and green space. In China there are several regional organizations with WWF China based in Beijing. Most of the regional and local branches of WWF China operate independently from each other which are largely due to the nature of China's government system. The NGOs that operate do so at the good graces of the CCP which would try to thwart any large, centralized national NGO network. This forces NGOs to operate in the decentralized, autonomous way that WWF China operates. WWF Shanghai has around 1000 members listed but really has only a few dozen who are active.³⁵ At the time of the Jiangwan incident they had not registered officially as a legal NGO so their current status was an illegal organization. Their activities are actually fairly limited to meeting together and bird watching at various places outside and around Shanghai and not the activism one would expect from such an so that's why they decided not to register.³⁶ Members mostly communicate on the WWF Shanghai's website to organize meetings and expeditions. By its nature, this ENGO is fairly unorganized in its internal structure. Again this is a product of the illegal nature of this organization, by organizing the internal structure of the ENGO it would risk the government attempting to clamp down on it. It has also prevented WWF Shanghai from increasing its links with other NGOs in the area to coordinate activities.

The Jiangwan marshlands were originally inhabited by an airport built for the Chinese military before occupying its role as an important green space in Shanghai. After many years of municipal growth for Shanghai and with the construction of other airbases, the airport became strategically irrelevant. During the 1970s, the area was turned into a marshland and control was given to the Shanghai municipal government. Due to the lack of human disturbance, the area was transformed into a haven for the city's wildlife. There are a large amount of flora and fauna that live there and it is a breeding ground for 90% of the

city's birds.³⁷ Due to the area's large size (about 1% of the city's area) and location within Shanghai, the municipal government wanted to transform it into a residential community. Considering its importance for the city's wildlife, especially its bird population, WWF Shanghai took the leadership in trying to stop this development from occurring.

The activists encountered many barriers to effectively reach out to others outside their organization for support. The first was to assemble a report on the environmental impact of destroying the marshlands. Very few scientists wished to put their name on anything which would indicate that they cooperated with an illegal organization.³⁸ Eventually, WWF Shanghai was able to get one of its own activists with a scientific background to write one. The NGO attempted to reach out to other organizations like the Grassroots Community as well as various student groups which were met with failure. Many of these groups thought their legal status would be threatened if they cooperated with an illegal group. Others were instructed by the municipal government specifically to take no action on this issue. The only groups who were willing to participate were those other web based groups like Shanghai Nomad Cub and the Common Wealth Management Project.³⁹

WWF Shanghai and their partners attempted to use mass media to attract attention to their fight to protect marshlands. They attempted to contact journalists but the municipal government put pressure on the news agencies leading to only a few stories to be produced on the issue. Eventually the media stopped reporting on their situation all together. The only other option for these groups was to use the internet in order to raise awareness. This strategy that WWF Shanghai was forced to use was actually fairly effective at getting their message out since a large number of Shanghai residents had access to the internet.⁴⁰ They used the forums on their partner groups like the Nomad Club and Common Wealth to report on the destruction. The amount of visitors being routed to the WWF Shanghai website kept rising so dramatically that the government had to shut down the website all together. Though the damage was done and public opinion had already started to turn against the government.

³⁵ Xie, 146.

³⁶ Ibid.

³⁷ Xie, 149.

³⁸ Ibid.

³⁹ Ibid.

⁴⁰ Ibid.

Because of the notoriety of the Jiangwan Marshland campaign after several months of online campaigning, more of Shanghai's citizens started to become involved in the project. Many different environmental experts were organized to offer their opinions on destruction of the marshlands at a level that was not possible before. An online petition against the construction was started and got the signatures from over 1000 people.⁴¹ It wasn't long before several members of the Shanghai Municipal People's Congress began to support the activist's aims. They thought that perhaps it was a good idea to possibly consider making Jiangwan a nature preserve. When the SMPC met for the next time all construction was stopped on the marshland. This was significant because "for the first time, the public, led by activists, scholars, and political representatives, had successfully influenced government policy." The public was able to lead to an outcome that wouldn't have happened through "institutionalized policy channels."⁴²

Despite the initial victory, there were still elements within the Shanghai municipal government who wished to develop the marshlands. First, the government targeted the scientists who supported the movement and were able to make them back down by pressuring them through their universities. The media was also more closely scrutinized to make sure that there was no more mention of the activists work. As the influence of the movement started to fade, the government conducted their own surveys to test public support over the marshland construction. They were able to produce an opinion outcome that was favorable for construction on the marshlands. This was done after a large public relations push on the side of the government who issued articles and news reports favorable to the construction in the media. Eventually, the government issued a new development plan for the marshlands which only preserved 200 square meters. WWF Shanghai was not able to meet its goals or objectives at preserving this critical green space within Shanghai.

Analysis and Policy Prescriptions

The PRC uses the system of NGO regulation as a way to control civil society within its borders. The process of recognition for NGOs is essential for this control to be effective since the government can basically decide which organizations it wants to

see operate. The existing NGOs are essentially beholden not to its members but rather to the government itself. "The Party monopolizes the legitimate channels of interest articulation and aggregation"⁴³ so their message has to agree with the party's. Since they decide if an organization is recognized or unrecognized, the organizations have a very limited arena where they can operate. If their issue is too controversial or political in nature then they will get shut down. NGOs literally have to live in a state of fear because any misstep might lead them to be de-recognized. This is why they were so averse to assisting WWF Shanghai since the government might act against them. De-recognition would mean that they would not be able to grow too large, advertise extensively, advocate for too many issues, or have many activities. Essentially, they would have to operate as a small, decentralized, and unorganized group much like WWF Shanghai did before they advocated for the Jiangwan marshlands. These NGOs "lack the organizational capacity to implement even medium-scale projects on their own."⁴⁴ They have to reach out and collaborate with others which is difficult if you are unrecognized because of the stigma attached to it. As an alternative, the government often tries to set up NGOs which have its leaders on the government payroll. They use the excuse that they want to eliminate "unproductive competition" between NGOs when what they really want to do is achieve control.⁴⁵ Just like business, the government wants to be able to control all areas of society in China so the most favorable outcome for the CCP is assured. One would ask why the government spends so much time suppressing organizations that are of a non-political nature. It is because these organizations "enhance individuals' capacity for collective political action."⁴⁶ They enable individuals to advocate for non-political issues which give them this experience for pushing for government change. By having all of these organizations providing this advocacy training to its population, in the future it can be directed towards political issues like voting and civil rights. Since political mobilization is one of the CCP's major fears, it is only natural that they clamp down on these NGOs.

⁴³ Guo, Gang. "Organizational Involvement and Political Participation in China," 460.

⁴⁴ Lu, Yiyi. "NGOs in China: Development Dynamics and Challenges," 93.

⁴⁵ Lu, 91.

⁴⁶ Guo, 457.

⁴¹ Xie, 152.

⁴² Ibid.

Another interesting phenomenon that this case study exposes was the use of the media and the internet to mobilize opposition to the project. Since the structure of WWF Shanghai is very loose and decentralized, the internet gave it the perfect medium to communicate with supporters. It was able to organize meetings of people to discuss the issue and come to a consensus on a plan of action. More importantly, it provided a medium of communication that was an alternative to the state run media. When the opposition to the marshlands got too threatening to the municipal government's plans, they ordered the media to stop reporting on it. This meant that the group's website and its partner's websites were the only way to inform the public about the marshland issue. With the media reports being skewed by the government the internet posed a contradiction to the official account of what was going on. The members of WWF Shanghai were able to, through their website, "offer alternative perspectives on events from those provided by other media."⁴⁷ Since most Chinese citizens are aware of the government's role in the press, they might be more inclined to believe what they see on the web from independent sources. Suddenly, the PRC's stranglehold on the media and official communication is loosened by the internet. WWF Shanghai even got a petition signed by over 1000 Shanghai residents which assisted in getting a temporary halt on the project. As was seen with the first case study, petitions are an effective and mostly legal tool for citizens to register their discontent. But why is the CCP so afraid of this activity which ultimately made them close down the website? It is what Howard Rheingold describes as the internet's "democratizing potential."⁴⁸ This potential comes from where the fast transfer of information between millions of individuals can lead to. If citizens can communicate instantaneously and organize themselves to advocate for environmental issues, why not for democratic rights as well? It is the same dilemma as the NGOs pose; their work and experience given onto citizens for non-political issues can easily be turned onto political issues. The internet threatens to "transform their more traditional forms of political behavior" into that which can directly oppose the government.⁴⁹ With the rising amount of Chinese citizens who have access to the internet this can be extremely dangerous for the regime.

⁴⁷ Rawnsley, 127.

⁴⁸ Rawnsley, 126.

⁴⁹ Ibid.

Shanghai is a perfect example of this potential since it has a very large internet savvy population.

This case study shows that even within a certain level of government in the PRC, there are a variety of interests and actions at work at one time. Certain members of the SMPC were sympathetic to the people's cause of protecting the marshlands. WWF Shanghai's petitions and scientific studies were able to convince the Congress that the marshlands needed more examination before it was built on. Though the council was able to effectively stop the construction, there were other interests in the mayor's office, the city planning board and other municipal institutions which wanted to see it continue. The higher officials in the government, by the system's nature, are more worried about developing the area and advancing their economic goals so they can get posted to higher positions by the central CCP. This shows that even in a city where the environment does place as a high priority, development still wins. It is economic progress that ultimately gets one promoted, not progress on protecting the environment. Because the interests of the different levels of the municipal government diverge with the more powerful ones in favor of development, ultimately the builders won.

The PRC needs to ease its restrictions on NGO registration and development in order to properly address the environmental concerns of its citizens. In a state such as China, there is no democratic citizen-government feedback loop for people to air their grievances and express their concerns. NGOs provide an alternative to that loop by providing a collective voice for issues of the population. By cutting that alternative out of the picture, citizens lack a mechanism to register their disagreements with government policy and mismanagement which could become a potentially explosive issue. By keeping the cap on civil society, China's goal of political stability could be undermined as environmental problems get even worse and increasingly dissatisfied people have no outlet for their concerns. The government has "limited NGO's usefulness as champions of interests and values different from that of the government."⁵⁰ By restricting NGOs, the PRC misses an opportunity to use them for their non-advocacy benefits as well. Plenty of NGOs do work in environmental cleanup, conservation activities, and environmental management which would go a long way to helping China cope with its environmental issues. Without the opening up of the registration system there will

⁵⁰ Lu, 93.

be few national networks of NGOs and little cooperation between them. This is required for any healthy society to flourish as China becomes more urbanized and its citizens look for these organizations to fill certain gaps.

At first, WWF Shanghai as an illegal NGO was able to work around many barriers that the government imposed with registration like cooperation with others, media and the internet. But in the end the government was able to halt its activities by cutting off its institutional links and outlets for organization. At this point in China, it would be extremely difficult for illegal NGOs to succeed in advocacy efforts because of the government can easily stop them from operating. The internet became a useful tool for WWF Shanghai in mobilizing supporters and spreading their message. It is difficult for the government to keep track of all traffic on the web which allowed even an illegal NGO to succeed for so long. The fact that they almost succeeded in getting the project completely stopped could provide a strategy for other illegal NGOs to use. Certainly for a legal NGO these benefits would accrue quicker with an established network of supporters and more range to operate due to its status. Despite WWF Shanghai's limited success, it ultimately failed because of their lack of recognition and the benefits afforded from that. In order for the NGO society to operate effectively and accurately reflect the people's interests, China will have to deregulate its NGO sector and allow for freer access to the internet.

Case Study #3: Environmental Activism from Beijing to the Nu River

Beijing, like Shanghai and many other cities in China has experienced extremely fast growth over the past couple decades. Also like many other Chinese cities, it has a very poor environmental protection record with serious air and water pollution problems.⁵¹ It wasn't until the 2000s when Beijing sought to improve its pollution problem culminating in a fierce campaign right before the 2008 Olympic Games. Unlike in other regions, Beijing has been fairly tolerant of NGOs operating within the city. The Ministry of Environmental Protection (MEP) has openly declared their support for many such institutions and has even cooperated with them in the past. It has even included their participation in

⁵¹ Xie, 86.

Environmental Impact Assessments (EIA) which are issued before building on projects commences. Part of this reason was an effort by the MEP to achieve more influence on the EIA process where it is often eclipsed by other economically oriented ministries.⁵² After 56 ENGO's issued a letter in 2005 which supported the MEP's growing influence on the EIA process, an informal partnership was established between these two groups. This ultimately led to the establishment of the regulation "Provisional Measures for Public Participation in Environmental Impact Assessment" in 2006 which was the "first legal document that enables the participation of the public in environmental policy-making process."⁵³

A large number of public ENGOs exist in Beijing and foster a community of organizations which are able to advocate for their interests. They have their roots in the 1990's with the establishment of the first ENGO, Friends of Nature, in 1993. An offshoot of that NGO, the Green Environmental Volunteers (GEV), is the ENGO that we will be studying in this case study. After the turn of the century the amount of ENGOs increased dramatically after feeding off of existing organizations and enjoying increasing support from the MEP. Student green organizations also became prominent which fed into larger organizations after graduation. ENGOs in Beijing have had a large degree of cooperation by "sharing information, organizing workshops, launching programs collectively to achieve stronger influence" etc.⁵⁴ The GEV in particular was one of the most prominent ENGOs in Beijing. They had almost 200 permanent members with a volunteer and advocacy network of over 30,000 people.⁵⁵ It is well connected and has led a number of campaigns with other NGOs before. Despite its loose organizational structure, its leadership of extremely motivated individuals enables it to be successful.

The GEV took a leadership role in opposing the decision to build a large scale damming project on the Nu River. The Nu river flows through Yunnan province into Myanmar and is one of the last large rivers to not be dammed in China. Surrounding it is a natural World Heritage Site where much of China's biodiversity lies with 79 rare or endangered species making their home

⁵² Xie, 87.

⁵³ Ibid.

⁵⁴ Xie, 90.

⁵⁵ Xie, 92.

there.⁵⁶ The project entailed building 13 hydro-electric power stations at various points on the river generating 3.64 million kilowatts which is more than ever the Three Gorges Dam.⁵⁷ Due to the environmental impact of the dam, which would disturb much of the wildlife and plant species in the area, there was quite a lot of opposition to the project. The MEP and a group of ENGOs were opposed to the project and tried to organize meetings and make the public aware of the downsides of this project. The National Development and Reform Commission (NDRC) was on the other side of the issue claiming that the benefits of this badly needed power generation outweigh the downsides. The controversy over the project was able to push the projected start date past 2003 giving the ENGOs a chance to step up their campaign against the dams.

NGO partnerships between the GEV and local Yunnan organizations were crucial in opposing the dam project. In Kunming, the capital of Yunnan, the ENGO Green Watershed was the primary organization in opposition to the project. Support locally was weak for Green Watershed because the provincial government cracked down on potential partnerships by restricting media reporting as well as by other means. Upon interviewing leaders of other NGOs, it was discovered that their lack of participation was due to concerns over their status which could be under threat if they oppose the government.⁵⁸ Due to this lack of support, Green Watershed was forced to turn to GEV, a Beijing based NGO in the hope of assistance. GEV was very willing to assist in advocating against such a potentially damaging project and was able to get several of its partners involved in the fight. Because restrictions on the ENGO culture in Beijing were relatively loose, many of the leaders of the different organizations in the city had very developed *guanxi*, or influential networks of influence.⁵⁹ In addition to the NGO sector, GEV's leaders were able to use their *guanxi* with the MEP to coordinate action.

Once these partnerships were made between regional and Beijing NGOs they started to act to preserve the watershed. The MEP attempted to get a serious EIA of the project but it was stonewalled by supporters within the government. Many in the MEP increasingly turned toward the ENGO sector

by providing them with updates and information in order to help them advocate strongly for the area's preservation in ways that they could not.⁶⁰ The GEV was then able to use this information to report this issue in the media and get scientists to support their environmental claims. Through both organizations, scientists in both Beijing, and locally in Yunnan were contacted to offer their opinion through both on the ground assessments and reports. The MEP could then turn around and say the original EIA of the site was biased because of the clear opposition of all these scientists without seeming like it was instigating a coalition against the government. The media was also able to publish many of these opinions in both public newspapers and, most significantly, the daily newspaper for the CCP. This newspaper had a real significant impact on policy makers which was a major achievement for the movement since their goal was not only to reach a public consensus against the building but that of the government as well. Eventually the GEV organized a trip for journalists to the area where they got to see the environmental as well as social impact of the project. It was a big concern that over 50,000 people would have to be moved from their homes as a result of the dams.⁶¹ The more media coverage the issue got the more support from the public, government, and other ENGOs the anti-building side received.

As public support continued to rise against the dam building, the actions of the ENGO coalition intensified in 2004. They conducted a series of photo exhibitions at several universities, in a supermarket, post office, and an office building showing the Nu river watershed. They even created a website called "Nu River Sentiment" where people could pledge their support for the region's preservation.⁶² More scientific studies were published and presented at conferences nationally and internationally. The coalition even got international support from NGOs further downriver from the dams. 80 NGOs in Thailand and Laos wrote letters protesting the project and sent them to the Chinese embassy in Thailand.⁶³ In the end, due to a combination of pressures, the government continued to push off the project with no definite start date for construction at present time.

⁵⁶ Xie, 97.

⁵⁷ Ibid.

⁵⁸ Xie, 100.

⁵⁹ Ibid.

⁶⁰ Xie, 101.

⁶¹ Xie, 103.

⁶² Ibid.

⁶³ Xie, 104.

Analysis and Policy Prescriptions

The success of the Green Environmental Volunteers was significant in both the size and scope of the Nu River dam project. The GEV was a legal NGO so therefore it had more room to operate than the illegal NGO we saw in the first case study, WWF Shanghai. The advantages afforded to it because of its status were obvious. It was allowed to use the internet and reach out to the public without being clamped down upon. Organizations were able to partner up with it without fear of reprisals from the government. It was even allowed to cooperate with the Environmental Protection Ministry which would have been forbidden if the GEV was an illegal group. Without these benefits the campaign probably would not have been able to attract as much public attention. The media was allowed to report on the issue as much as it wanted as long as it didn't openly challenge the government. Journalists were not as afraid as approaching the NGO and the government did not stop news stories from being published.

The GEV was permitted to operate as it did partially because of the lack of crackdown from the Beijing government. Even though it had legal status, the government might have been able to try to shut it down by other means but it didn't. The Nu river dam project was of the utmost national significance and posed an opportunity to greatly increase China's power generation capacity. The potential of this project was more than that of the Three Gorges Dam, arguably China's most important infrastructure project. Part of the reason that this crackdown didn't happen was because of the central government's new priorities for environmental protection. It was discovered that "environmental pollution cost China 511.8 billion Yuan (USD64 billion) in economic losses, accounting for 3.05 per cent of the country's GDP in 2004."⁶⁴ This only quantifies what China has known for years which were that these problems were hurting China's economic growth and the problem was getting worse. If China was to maintain its high growth levels it had to deal with these problems. It also had to tackle many of its problems due to its expected accession to the WTO and in preparation for the 2008 Olympic Games. Because these organizations were located within Beijing, central government and the city's policies

often overlapped. Since environmental protection was acceptable to the central government up until a point, these NGOs were permitted to operate extensively.

The establishment of the Ministry of Environmental Protection is a step in the right direct yet shows some unsettling aspects of Chinese environmental policy. The Ministry was a relatively new creation when it was created as the State Environmental Protection Administration in 1998 when environmental protection in China became a larger priority.⁶⁵ As seen in this case study, it began to partner up with NGOs in the early to mid-2000s to help advance its goals. Part of the reason that this cooperation with civil society was needed was because it "lacks the money, staff, and power to override local leaders and to shut down polluters."⁶⁶ Unable to achieve its goals itself because of the lack of resources, it had to turn to these sectors. The information that it provided to GEV was crucial in allowing scientists to compile reports claiming the marshland construction was damaging. In giving them support, the MEP was able to use the GEV to force the opinion of the public against the opinion of more powerful ministries like the NDRC. Elevating environmental protection to ministry level should not give the impression that it has achieved the same priority level as the NDRC which has equal status. The MEV is handicapped at the start because "environmental protection often loses to the more attractive goal of economic development."⁶⁷ With economic growth still remaining the top priority for the government, ministries like the NDRC will always have more resources and capabilities than the MEV. In a way this will continue to strengthen the NGO community since the state will have to continue to turn to "non-state actors to carry some of the environmental protection burden."⁶⁸ Because of this increasing role of ENGOs in environmental protection, it leads to new regulations passed to legitimize public participation in EIAs. The MEP's support of the ENGO community has been essential to its growth and to its protections efforts.

This Nu River damming issue exposes the stark difference between different agencies within the central government. The MEP has shown a

⁶⁴ Huang, Xibing. "Environmental issues and policy priorities in China: a content analysis of government documents."

⁶⁵ Schwartz, Jonathan. "Environmental NGOs in China: Roles and Limits."

⁶⁶ Liu, Jiangguo. "Revolutionizing China's Environmental Protection

⁶⁷ Schwartz

⁶⁸ Ibid.

remarkable amount of independence from the goals of economic policy planners within the PRC. It has been able to raise environmental concerns up to a point where their opinion is recognized as a legitimate consideration in the construction process. Much of the time this opinion is ignored but it is still significant progress since it is the starting point for further advancements. There is still a long way to go before it reaches the power levels of the NDRC who was starkly opposed to the project. The bureaucratic power struggle between the two ministries shows that the supposedly monolithic power structure of the Chinese government is not even true at the central level. Both agencies were ultimately engaged in a fight to convince policy makers that their position is the most legitimate. In this case, because public opinion was so strong because of the environmental damage, the “ecological impact” vs. “growth of the local economy and the increase in national energy” argument landed on the side of the environment.⁶⁹ Unfortunately this is a rarity because more often the side of the NDRC wins out making this victory more significant.

The PRC needs to strengthen its institutional capacity for environmental enforcement in order for the environment to be properly protected. The creation of the ministry is significant, but it still does not mean that its goals will be met. The government needs to give the MEP the capabilities to enforce environmental regulations and review projects more closely. This will not happen unless a number of changes happen to the ministry, including raising the numbers of employment from its current number which is one-eighth the size of the US Environmental Protection Agency.⁷⁰ Previously it was mentioned that this lack of support has strengthened the ENGO community by forcing the MEP to turn to them for assistance but a strengthening would also help the community. By giving the MEP the resources to accomplish its mission, it would strengthen the causes of these ENGOs by giving their positions more legitimacy from a branch of the government. The relationship between the ENGOs and MEP has already been established so it would be unlikely that it would be diminished just because one side’s power capabilities have been enhanced especially since they would be working in tandem. ENGOs need to continue to exploit the differences between branches of the government to help them succeed.

⁶⁹ Xie, 102.

⁷⁰ Liu

By combining this with favorable public opinion for preserving the Nu River watershed, the GEV was able to preserve the region for the time being.

Conclusion

The three case studies in this paper reveal several interesting developments in how the government and NGOs interact in the realm of the environment. One of the most significant is how different levels of government as well as different institutions inside those levels have such a variety of interests and methods for dealing with environmental activism. In our first case study we see a lower level of government (local) ignoring the regulations of the higher level of government (municipal) in order to build on park lands. The local level of government was able to influence the bureaucracy enough that they were willing to ignore the regulations as well. One would expect that rules coming from a higher level of government to a lower level of government in an authoritarian, one-party state to be followed no matter what but that is clearly not the case. Among other factors, once the higher levels of government was informed about this lack of adherence they put pressure on the local government to shut the project down. In the second case study at the municipal level of Shanghai there was a stark difference not between levels of government but among the same level of government. The Shanghai Municipal People’s Congress actually was able to delay the project for a time. They thought the project should be reevaluated before construction proceeded against the wishes of the higher officials within the city government. Ultimately the wishes of these higher officials were able to win out since the process went along. In the last case study there was a difference between two branches of the central government between the Ministry of Environmental Protection and the National Development and Reform Commission. The MEP wanted the Nu River project stopped and the NDRC wanted it to continue with both ministries trying to influence policy makers. These branches of the bureaucracy were literally competing for influence in order to advance their respective goals. Though the Chinese government operates in a one-party state the interests of the top do not necessarily lead to the desired outcomes at the bottom. This is because as we have seen with the case studies, every branch and level of government has goals and interests that

they try to serve and they vary enough that differing outcomes will manifest.

Activists have tried a number of strategies to overcome barriers that the PRC has placed to block their activism. Shen, an individual in the first case study was able to motivate his neighbors to activate for a purely local concern. He was able to use the actual laws of the state which forbade construction to validate his position. This made it difficult for government officials to argue with him or shut his protests down because they were using actual laws of the state to “authenticate” his position. He framed their protests as simply protecting the resident’s environmental rights which was extremely effective. When their interests were ignored, on two separate occasions Shen ordered that the protesters take down first the wall, and then the scaffolding at the site in an orderly manner. It was important that they did not destroy anything and that they did this on the basis of the law because if not, this protest would not have been tolerated. He was able to make the government afraid of their potential violence but not of any outward real violence which would have gotten them imprisoned. In Beijing, legal NGOs like the Green Environmental Volunteers had more leeway to operate. They were able to form direct partnerships with the MEP which was a district branch of the government. The MEP was even able to send information regarding the project enabling them to have scientists rally against the Nu River dam. Though rare, these partnerships with supporters in government are essential for NGOs to exploit. Since they are legitimate voices in policy making circles, they are able to influence decisions. In the case of Beijing, the MEP had enough strength to make a difference in that advocacy situation. The GEV was able to raise public opinion through the media and the internet and let the public know how destructive the project would be. Because the MEP coupled with NGOs who were able to get public opinion on their side, this greatly enhanced their power so that they could compete with the more powerful ministries like the NDRC. Another advantage that the GEV had in this situation was the ability to form connections with other NGOs. Since the zeal for environmental protection was so high at the time and Beijing was going through a period of relative loosening, this was possible. The GEV was able to use other Beijing NGOs to assist them at building public opinion through their *guanxi*, or personal connections. They were even able to assist local NGOs who exist in more unfavorable climates to

advocate for the construction to not proceed.

Forming these networks of cooperation is necessary for effective national opposition to environmental projects to succeed.

While there were many strategies that were effective at advancing the environmentalist’s goals there were also many others that were met with mixed results. The efforts of WWF Shanghai were the most prevalent in this case. They were an illegal NGO that never applied for recognition because they never needed to. When the issue for the protection of the Jiangwan marshlands became apparent, they attempted to advocate for its protection but instead were met with a clampdown from the government. They did attempt to use the media and the internet as tools for promoting their issues. Shanghai as well as many other parts of China is increasingly connected to the internet giving them access to more information. Because of this, WWF Shanghai, an illegal NGO was able to be successful for a short time even after the government stopped media reporting on the issue. They could organize meetings to come to a scientific consensus and also get petitions signed by other citizens. Their online efforts were almost effective when the SMPC ordered a stop on their construction efforts. At the same time the internet offers many advantages for the green movement, it also gives them several vulnerabilities. The government was able to shut down WWF Shanghai’s internet outlet and succeed at getting their project built because public opinion started to fade. With the GEV environmental activism online was permitted because they were a legal NGO but that is not always guaranteed to be the case if the government feels threatened. Nevertheless, the use of the internet is an effective tool for both illegal and legal NGOs. Ultimately if NGOs want to succeed at their activism in the long term, they will need this legal status to function effectively. The politicization of environmental concerns was also a big issue as was seen in the first case study. The government tried to drag the activism into the political realm as Shen tried to keep it nonpolitical. They almost succeeded at stopping his movement but because of Shen’s persistence at keeping the issue centered on the environment, they were able to succeed. For the time being activists will have to make sure that their activism is kept in this non-political realm until there is a widespread political change within the PRC.

The PRC needs to embrace a number of reforms in order for it to better cope with environmental activism. Firstly, it needs to ensure a

greater rule of law within the bureaucracy and government. Regulations should be followed no matter what politicians and party members try to say to influence it. This is the only way that environmental regulations will be ensured to be effective at all levels and not be controlled by special interests. This move towards greater rule of law has other benefits for the PRC like greater efficiency and it seems that it is moving generally in that direction. The party would still ultimately be able to write what regulations which would be followed so nothing too far outside their interests would be done. The PRC also needs to lift the restrictions that it places on NGO registration in order for civil society to flourish and the interests of the people be properly reflected. Understandably these NGOs represent a threat to the regime since they give them activist training. But, in order to assist the state in identifying polluters and without the collaboration from NGOs to protect China's environment, the state's environmental needs will be difficult to meet. The government also needs to strengthen the MEP to make it able to compete with other ministries. Of course, economic interests will trump environmental ones for a long time. More money, resources and staff will still go far in ensuring that the MEP has more of an ability to articulate its interests. Many of these reforms would be beneficial, but China is only likely to implement them in a limited manner in the short term.

Ultimately what is necessary for China to have truly effective environmental activism would be an open political system and competitive democracy. Since this will not happen anytime soon, China needs to implement several of these reforms in order to effectively deal with environmental concerns of its citizens. Like mentioned before, environmental problems have explosive potential if they continue to deteriorate which represent "destabilizing factors" for the government. This is why the PRC has decided to take up environmental protection as one of its top concerns. Though commendable, they have not gone far enough in ensuring this protection leading to so much activism. These activists have met much resistance and encountered many failures but at the same time many successes as well. Jonathan Schwartz writes that "China's environmental non-governmental organization community-arguably the most advanced manifestation of Chinese civil society." In this role, environmental activists have the potential to define how civil society as a whole grows and changes in China into the future.

Bibliography

- "A Lot to be Angry About." *The Economist*. 3 May 2008, Print.
- Barboza, David. "China Leader Encourages Criticism of Government." *The New York Times* 26 Jan. 2011, Print.
- Guo, Gang. "Organizational Involvement and Political Participation in China." *Comparative Political Studies*. 40.4 (2007): 457-482. Print.
- Ho, Peter. "Who Owns China's Land? Policies, Property Rights and Deliberate Institutional Ambiguity." *China Quarterly*. 166 (2001): 394-421. Print
- Huang, Xibing. "Environmental issues and policy priorities in China: a content analysis of government documents." *China: An International Journal*. 8.2 (2010): 220-240. Print.
- Liu, Jiangguo. "Revolutionizing China's Environmental Protection." *Science*. 319.5859 (2008): p.37-38
- Lu, Yiyi. "NGOs in China: Development Dynamics and Challenges." *China's Opening Society*. Ed.. Zheng Yongnian. Milton Park, UK: Routledge, 2008. Print.
- Peerenboom, Randall. " A Government of Laws: democracy, rule of law and administrative law reform in the PRC." *Journal of Contemporary China*. 12.34 (2003): 45-67. Print
- People's Republic of China. *Constitution of the People's Republic of China*. Beijing: 1982. Print. Jan 2011, Print
- "Reassessing George Bush." *The Economist*. 11 Nov. 2010, Print.
- Rawnsley, Gary D.. "Media, Internet and Governance." *China's Opening Society*. Ed.. Zheng Yongnian. Milton Park: Routledge, 2008. Print.
- Schwartz, Jonathan. "Environmental NGOs in China: Roles and Limits." *Pacific Affairs*. 77.1 (2008): p.28-51. Print.

Tong, Jingrong. "Press self-censorship in China: A case study in the transformation of discourse." *Discourse Society*. 20.5 (2009): 593-612. Print.

Xie, Lei. "Environmental Activism in China." New York: Routledge, 2009. Print.

Zhu, Jiangang. "Not against the state, just protecting residents' interests." *China's Embedded Activism*. Ed. Peter Ho. New York: Routledge, 2008. Print.

Iran's Nuclear Program

Johnson Lin

I. Introduction

The potential Iranian nuclear weapons program is an issue of utmost concern to the international community. Iran has been deemed by the U.S. State Department as a state sponsor of terror, and it has the dangerous potential to use its nuclear program to construct nuclear weapons. Iranian leaders have consistently declared Iran's right to pursue peaceful nuclear technologies, but they have not fully adhered to the international inspections and regulations that are necessary to ensure its peaceful nature. While there isn't irrefutable evidence linking Iran's nuclear advancement to nuclear weapons, many signs seem to point in that direction.

The problem of Iran's nuclear program fits perfectly within the realist theory of international relations. Despite the existence of the United Nations and other multinational organizations, realism argues that states exist in an anarchic environment. Uncertainty about the capabilities and intentions of Iran's nuclear program creates serious security dilemmas for the international community, particularly the U.S. and Israel. The international community has acted to thwart Iran's suspect nuclear weapons program. Some of the options available can be successful and others are doomed to fail. Understanding the history of Iran's nuclear program is necessary for understanding how the international community ought deal with this potentially deadly problem.

II. History of Iran's nuclear program

Iran's nuclear program began in 1957 through the United States "Atoms for Peace" program. Under President Dwight D. Eisenhower, the program promoted the development of safe and peaceful nuclear technologies. It strengthened the alliance between the two, which was vital to U.S. national security interests because of the Cold War and Iran's strategic location on the border between the Soviet Union and the Middle East. President Eisenhower sought to increase military, economic, and civilian assistance to gain a greater strategic position with Iran. The aid to Iran's nuclear program grew from U.S. national security interests, an example of international realism. Under Atoms

for Peace, the United States agreed to lease "Iran up to 13.2 pounds of low-enriched uranium (LEU) for research purposes" and also supplied a five-megawatt reactor for the production of electricity.¹ In return for U.S. nuclear assistance, Iran signed the Nuclear Non-Proliferation Treaty (NPT) on July 1, 1968, the first day the treaty was open for signature.

The NPT originated from the international community's growing awareness of the inherent dangers of the proliferation of nuclear arms. After witnessing the destructiveness of nuclear weapons against Japan during World War II, many believed an increase in the number of countries with nuclear weapons threatened global security. The treaty served international security interests by advancing nuclear non-proliferation, disarming and dismantling existing nuclear weapons, and advocating for peaceful nuclear development.² Iran, having just initiated a meaningful nuclear program, signed the treaty as a non-nuclear signatory, agreeing not to acquire or manufacture nuclear weapons, to accept safeguards that verified nonproliferation obligations and to negotiate in good faith for the cessation of the nuclear arms race. These conditions were supposed to bar Iran, and other non-nuclear states, from developing nuclear weapons. In return for its agreement to these provisions, under Article IV of the treaty, Iran gained recognition of its "inalienable right...to develop research, production and use of nuclear energy for peaceful purposes" in accordance to articles I and II.³ Iran's signing of the NPT publicly signified its intention to abstain from developing nuclear weapons.

With a commitment to peaceful and responsible nuclear development, Iran was able to acquire nuclear assistance from Western states such as France and Germany throughout the 1970s. The nuclear program was a priority of Shah Reza Pahlavi, Iran's leader, because he believed that the exploration and development of alternative forms

¹ Greg Bruno, "Iran's Nuclear Program," *Council on Foreign Relations* (March 10, 2010)

² Paul Lettow, "Strengthening the Nuclear Nonproliferation Regime," *Council on Foreign Relations* 54 (April 2010)

³ Lettow, "Strengthening the Nuclear Nonproliferation Regime"

of energy production was necessary for the Iranian economy given the economic uncertainty resulting from regional wars that disrupted Iran's oil production and from predictions of looming energy shortfalls.⁴ Instability in the Middle East in the 1970s pushed Iran towards developing nuclear energy, a means of meeting its energy demands beyond oil. In March 1974, Iran announced its plans to produce 23,000 megawatts of electricity from nuclear power. This ambitious goal would have been able to meet a great deal of Iran's energy needs, but the amount of nuclear energy necessary to do so was also enough to produce about 600-700 nuclear warheads.⁵ Iran's growing ambitions about the possibilities of a large nuclear program began to rattle its partners around the world.

Despite the Shah's assurance that his nuclear program had only peaceful intentions, foreign governments increasingly became reluctant to aid Iranian nuclear development. A United States special national intelligence estimate asserted, "The ambitions of the Shah could lead Iran to pursue nuclear weapons, especially in the shadow of India's nuclear test in May 1974."⁶ Furthermore, in an interview with the French newspaper *Le Monde*, the Shah was quoted as saying that Iran would, "sooner than is believed" be "in possession of a nuclear bomb," and Assadollah Alam, the Shah's court minister, claimed on several occasions that the Shah wanted a nuclear bomb but found it expedient to adamantly deny it at the time.⁷ These indications that Iran's nuclear ambitions was leaning towards more than merely energy production pushed the U.S. towards taking a stiffer approach against Iran.

By the time of Iran's Islamic Revolution in the late 1970s, France and Germany began withdrawing their support of Iran's nuclear program. The chaos resulting from the revolution, primarily the hostage crisis and the regime change to a hostile Islamic republic, created a dangerous situation that Western governments believed would only be exacerbated with the inclusion of nuclear technologies. During this period, Iran's nuclear program slowed dramatically. The deterioration of the program was the result of a combination of factors: the absence of foreign assistance, Supreme Leader Ayatollah Khomeini's opposition to nuclear

technology, the mass exodus of nuclear scientists, and Israel's destruction of Iraq's nuclear facility at Osirak, which removed an immediate threat to Iran.⁸ However, this break in Iran's nuclear development was ephemeral. It ended with the death of Ayatollah Khomeini in 1989 and the subsequent ascension of Ayatollah Khamenei. This change of Iran's leadership changed the trajectory of Iran's nuclear desires.

Ayatollah Khamenei had much more of a positive view of nuclear power than his predecessor, and initiated plans to rebuild the nuclear program in Tehran. In the 1991 Gulf War against Iraq, coalition forces, of which Iran was not a participant, discovered a clandestine Iraqi nuclear program. With the increase of a Western presence in the Persian Gulf and the real possibility of Iraqi nuclear armament, Iran perceived increasing security threats. With a neighboring rival having almost developed nuclear weapons, and Western powers exerting influence in the region, Iran believed that it could no longer depend on its traditional non-nuclear defenses.⁹ Following the tenets of realism, insecurity pushed Iran to further develop an indigenous nuclear program.

In addition to defense, nuclear weapons would also be useful for Iranian attainment of regional hegemony. Iran remembers the glories of the ancient Persian Empire, which was decimated by raiding Arabs, who were spreading Islam across the region. Iranians beam with nationalism and see themselves as separate and superior to neighboring Arabs. This is evident in Iran's adherence to Shi'a Islam as opposed to the Sunni sect popular among Arabs. Following the fall of the Soviet Union, the Middle East was no longer a battleground between superpowers. New regional leaders could emerge, and a functional Iranian nuclear weapon would cement Iran's superiority over its Arab rivals and leadership in the region. In order to construct nuclear weapons, Iran has received assistance from non-Western states such as Russia, China, Pakistan, and North Korea.¹⁰

As Iran accelerated its development of nuclear technologies in the 1990s, it began moving astray from its commitments as a signatory to the Nuclear Non-Proliferation Treaty. With insufficient oversight from the International Atomic Energy Agency (IAEA), Iran developed "a vast network of uranium mines, enrichment plants, conversion sites,

⁴ Bruno, "Iran's Nuclear Program"

⁵ Abbas Milani, "The Shah's Atomic Dreams," *Foreign Policy* (December 29, 2010)

⁶ Bruno, "Iran's Nuclear Program"

⁷ Milani, "The Shah's Atomic Dreams"

⁸ Ibid.

⁹ Ibid.

¹⁰ Ibid.

and research reactors” and constructed other nuclear facilities, to supply domestically produced nuclear fuels. This immense network of uranium production now allows Iran to produce an estimated 2.77 kg of low-enriched uranium per day.¹¹ Due to Iranian duplicity, diplomatic stalling, and unwillingness from the international community to enforce the conditions of the NPT in a meaningful way, Iran has been uninhibited in its enrichment of uranium “on a scale that cannot plausibly be meant to produce fuel for a nuclear power reactor but could be used to produce the fissile material for a nuclear weapon.”¹² In addition to the large scale of Iran’s nuclear development, the nuclear program has been kept extremely secretive. This secrecy seems indicative of intent to hide aspects of a nuclear program that are unacceptable to the international community. If it were completely peaceful, it could likely gain foreign assistance in return for international oversight. Iran consistent refusal of international oversight, despite the possible benefits of it, signals that its nuclear ambitions are not completely peaceful.

Despite speculations of Iran’s intention to construct nuclear weapons, there isn’t any conclusive evidence proving it. There is uncertainty about its intentions because the “existing means of detecting, determining, and enforcing violations of safeguard obligations, and the will to do so, are inadequate for the task.”¹³ The IAEA is incapable of decisively determining Iranian intentions. While unlikely, in light of the lack of evidence, it is still a possibility that Iran’s nuclear technology will pursue strictly peaceful purposes. Iranian antagonism towards international interference in its nuclear development may be more indicative of its distrust of the international community rather than of its intentions of secretly constructing nuclear weapons.

Iran’s current nuclear technology and weapons technology necessary for the transportation of possible nuclear warheads is largely based on foreign assistance, particularly from North Korea. North Korean technicians and military experts have supported Iran’s path to achieve nuclear bomb capability and have aided in increasing the range of its ballistic missiles. A majority of Iran’s nuclear centrifuges are primarily based on North Korean designs, which are more technically advanced and

greater in size than those indigenous to Iran.¹⁴ This relationship between Iran and North Korea is extremely alarming given North Korea’s status as the only state to have officially withdrawn from the NPT, and to have developed nuclear weapons after signing the treaty. Iran’s dependence on North Korea may be a signal of its intentions to follow a similar path of nuclear development.

It is also believed that Iran gained some nuclear assistance from Pakistan. Abdul Qadeer Khan, the father of Pakistan’s nuclear program, had operated an international nuclear black market in the 1990s. In 2004, he confessed to this network of proliferation, and is believed to have contacted and sold centrifuges and blueprints to Iran in 1994. Following his confession, Pakistan took steps to neutralize Khan’s network to safeguard its nuclear weapons and technologies. However, there is still concern that parts of Khan’s black market network could still be open, and that Iran is the most active customer in the nuclear black market. Iran’s access to nuclear materials beyond the supervision of the international community is a source of concern.¹⁵

In recent years, the international community has become increasingly alarmed at the potential of a nuclear-armed Iran. Western intelligence agencies have attained large amounts of information that “provide snapshots of an extensive and sophisticated weapons program whose goal is to produce a uranium implosion device.”¹⁶ In response to criticisms and accusations of its nuclear ambitions, Iran consistently declares its sovereign right to pursue nuclear technology. This pursuit of nuclear technology has gotten alarmingly close to nuclear capability, and Israeli intelligence estimates that with the current rate of Iran’s development, it is about three years away from developing a nuclear weapon.¹⁷ Many believe that Iranian nuclear weapons are imminent. While a nuclear-armed Iran would be a threat to the entire international community, Israel would face the

¹⁴ Global Security Newswire, “Iran Seeks Nuclear Scientists, Former Diplomat Asserts” (December 23, 2010)

¹⁵ David Montero, “Report: Pakistani scientist A.Q. Khan aided Iran,” *Christian Science Monitor* (May 4, 2007)

¹⁶ Mark Fitzpatrick, “Iran’s Nuclear, Chemical and Biological Capabilities – Executive Summary,” *International Institute for Strategic Studies* (February 3, 2011)

¹⁷ Arthur Bright, “Iran nuclear program delayed, says Israeli minister,” *Christian Science Monitor* (December 30, 2010)

¹¹ Ibid.

¹² Lettow, “Strengthening the Nuclear Nonproliferation Regime”

¹³ Ibid.

greatest security threats from Iran, aptly expressed in current Iranian President Ahmadinejad's persistent anti-Zionist rhetoric. In 2006, during the war between Israel and Hezbollah in Lebanon, Ahmadinejad said in regards to the conflict, "...the main solution is for the elimination of the Zionist regime" because Israel "is an illegitimate regime, there is no legal basis for its existence."¹⁸

Along with President Ahmadinejad's harsh rhetoric, Iran's weapons capabilities have greatly increased in recent years, and now pose a serious threat to Israel's national security. Its advanced weapons are seen as a precursor to greater military ambitions that include nuclear weapons. In 2008, Iran test-fired nine missiles, including an advanced version of the Shahab-3, a medium-range ballistic missile. It reportedly has a range of up to 2,000 kilometers, and can be armed with a conventional high-explosive warhead, a chemical warhead, submunitions warhead, or a nuclear warhead.¹⁹ The Shahab-3 missiles can theoretically carry possible Iranian nuclear warheads to Israel, placing Israel in a very precarious situation. As such, Israel is often the most assertive member of the international community in calls for actions to stop Iranian nuclear development. Regardless of whether or not there is international support, Israel may act unilaterally to stop Iran's nuclear program.

Under the assumption that Iran is in fact developing nuclear weapons, the international community has sought to implement means of disrupting Iran's nuclear advancement. Multilateral and unilateral, specifically Israeli, approaches have been discussed. Some efforts, such as sanctions and covert operations, have already been implemented. Others, such as preventive strike and regime change, are viable options in the near future. These options approach the problem from different perspectives and are not mutually exclusive. Regardless of which options are taken against Iran, success must be measured not only in whether Iran is able to construct nuclear weapons, but also cost to the quality of life of the Iranian population, and the potential of escalation to war.

III. Option 1 - Accept proliferation

Although many believe that the dangers posed by allowing Iran to continue its nuclear program demand international actions preventing its construction of nuclear weapons, the potential dangers may be overblown. The fact that there isn't any conclusive evidence about the purpose of Iran's nuclear program must be emphasized. Also, recent events in the Middle East and within Iran have triggered a de-emphasis on the nuclear program as Iranian leaders seek to satisfy economic demands by its public. Moreover, even if Iran were to develop nuclear weapons, it may not be any more threatening to the international community than it currently is. The option of inaction, and thereby allowing Iran to continue its nuclear program with the possibility of its constructing a nuclear weapon, is one that needs to be seriously considered by the international community.

With the unrest across the Middle East and rising inflation in Iran, Iran's nuclear program is not currently a priority for its leaders. 2011 has been a year in which the Iranian regime has placed greater emphasis on its domestic agenda because of rising inflation resulting from recent cuts to government subsidies on gasoline, fuel, and bread. With public outcries against its damped economy, developing nuclear technology has become less urgent for the regime. Inflation is "not at 10 percent but perceived by average Iranians to be 50 percent."²⁰ Sanctions on Iran have restricted its economy, forcing it to divert resources and attention away from producing a nuclear arsenal and towards dealing with its economy. A de-emphasized nuclear program means that immediate forceful action is neither required nor prudent. The pressures on its economy may force Iran to work with foreign states, opening the door to solving the nuclear problem diplomatically. Iran's nuclear advancement does not appear to be a threat in the immediate future.

On the other hand, even if de-emphasized, Iran's nuclear program is still a serious threat in the long run. If the international community allows Iran to continue its nuclear program and it successfully produces nuclear weapons, the international community fears that Iran will have increased coercive influence over its neighboring states. These fears may be unfounded since

¹⁸ Sean Yoong, "Ahmadinejad: Destroy Israel, End Crisis," *Washington Post* (August 3, 2006)

¹⁹ BBC News, "Iran sends missile test warning" (July 9, 2008)

²⁰ Scott Peterson, "New Year priorities: Tehran focused on turmoil at home, not nuclear program," *Christian Science Monitor* (December 31, 2010)

coercion backed by threats of nuclear weapons historically have not been more successful than coercion without such threats. For example, the United States' nuclear arsenal did not enhance the effectiveness of its demands of North Korea to release the USS Pueblo in 1968, Israeli threats against Syria prior to the 1982 Lebanon War were unsuccessful, and British threats against Argentina over the Falkland Islands in 1982 failed despite its possession of nuclear arms.²¹ One possible reason for the ineffectiveness of nuclear-backed threats is that the state threatened understands the unlikelihood of the actual use of nuclear weapons. Deterrents are ineffective unless the opposition believes that the threat is real and imminent. Any possible use of nuclear weapons would universally be condemned by the international community and would lead to serious repercussions, thereby neutralizing the threat. Following historical trends, Iran's possible nuclear weapons would not increase its coercive influence over regional neighbors.

Another fear of a nuclearized Iran is its potential to unleash a nuclear arms race in the Middle East. Secretary of State Clinton and Vice President Biden have both warned "Iran's actions risk sparking a nuclear arms race in the Middle East."²² While this is a possibility, it is also an overstated danger. There are only a handful of countries in the region, such as Libya, Syria, and Saudi Arabia, which have the technological capacity and the economic wherewithal to develop a meaningful nuclear program. Some of these states have previously had nuclear programs but have since given them up. Even if nuclear arms proliferated in the region, the real danger is not necessarily nuclear proliferation, but from a nuclear conflict.²³ The dangers of nuclear arms are not necessarily inherent, but result from their use.

Although Israel would be the most likely target of Iranian nuclear weapons, history indicates that nuclear proliferation has the effect of deterring nuclear wars rather than provoking them. Nuclear states may still conflict with one another, but nuclear weapons have never been brought into the conflict. For example, during the Cold War, neither the U.S. nor the U.S.S.R. used their massive

stash of nuclear weapons. In the current state of Indian-Pakistani relations, the two have not used nuclear arms despite mutual animosities. While these states conflict with one another, both sides understand the dire consequences of a nuclear attack. Pax Atomica, the theory of mutually assured destruction, deters the use of nuclear weapons. It is unlikely that nuclear war between Iran and Israel would occur even if Iran were to develop nuclear arms. Nuclear war between the two would be too dangerous and the notion of mutually assured destruction would encourage greater prudence in their bilateral relations. Iranian leaders understand that if they were to use nuclear arms, it would be reciprocated leading to massive destruction on both sides. This diminishes the likelihood of a nuclear conflict.

Although it is unlikely that Iran will use any nuclear weapons it possibly develops, a major fear is that the weapons will fall into the hands of terrorists. The proliferation of nuclear arms may spread to Al-Qaeda, Hezbollah, or other groups. In 1998, the U.S. Justice Department charged that Osama bin Laden called for "al-Qaeda to put aside its differences with Shi'ite Muslim terrorist organizations, including the government of Iran and its affiliated terrorist group Hezbollah, to cooperate against their perceived common enemy, the United States and its allies." Iran has indirectly supplied al-Qaeda and its affiliated groups with resources in their mutual war against the United States. However, it is unlikely that Iran would directly supply al-Qaeda with a nuclear weapon. The two share a common enemy, but also have divergent ideologies. Iran promotes Shi'a fundamentalist movements in the Middle East, while al-Qaeda is a Sunni jihadist movement. Iran would much more likely share its nuclear weapons with Hezbollah, a terrorist organization essentially created by Iran's Revolutionary Guards. Hezbollah has caused several conflicts with Israel since its founding in the 1980s and Iran has several weapons smuggling routes to the group that can be used to smuggle nuclear warheads to Hezbollah. Neither al-Qaeda nor Hezbollah has demonstrated reluctance in the possibility of using nuclear weapons.

The potential of terrorists groups using Iranian-developed nuclear weapons is a very deadly one, and is the most dangerous aspect of accepting proliferation. This may be the best argument against allowing Iran to continue its nuclear program. However dangerous this threat is, the possibility of it actually occurring is low when considering the reasons why Iran wants nuclear

²¹ Matthew Fuhrmann & Todd S. Sechser, "Would a nuclear-armed Iran really be so dangerous?" *Christian Science Monitor* (January 12, 2011)

²² BBC News, "US leader Biden says Iran may spark nuclear arms race" (May 6, 2010)

²³ Stephen M. Walt, "Iran, arms races, and war," *Foreign Policy* (October 1, 2009)

weapons. If the construction of nuclear weapons were key to Iran's regional hegemony, terrorist use of those weapons would surely diminish Iran's leadership in the region and would instead solidify Iran's status as a dangerous pariah state. The use of its nuclear weapons would be counterproductive to Iran's interests, and it would be in its best interest not to allow terrorist organizations to use it.²⁴

An Iranian nuclear weapons program is more likely to be used for defensive rather than offensive purposes. One of the key reasons for Iranian nuclear ambitions is perceived threats in the region. With the 2011 NATO air strikes in Libya, some see nuclear weapons as the only useful deterrent against foreign intervention. The North Korean regime has stated "Libya's dismantling of its nuclear weapons program has made it vulnerable to military intervention by the West." Halting a nuclear program is seen as a bait and switch approach that left Libya defenseless against consequent Western attacks.²⁵ Iran's regime, with close nuclear ties to North Korea, likely has a common understanding of the use of nuclear weapons as a deterrent against the West. With an increasing foreign presence in the Middle East, Iran sees its nuclear program as the only means of providing defense against perceived hostilities.

The option of accepting proliferation may not be as dangerous as is often assumed. While proliferation of nuclear weapons is dangerous, a nuclear-armed Iran will not likely be any more dangerous than any other nuclear-armed state. Based on historical precedents, Iran will not have any more coercive influence over its neighbors, nor will it start a nuclear conflict. It is possible that a nuclearized Iran would ignore historical precedent and launch a nuclear war, but in reality, Iran's leaders should be too rational to actually do so. The only serious threat of Iranian nuclear weapons is that terrorist organizations may procure and use them. However, this would run counter to Iran's interests. Instead, nuclear weapons will most likely be used for defenses against Western threats. Allowing Iran to construct a nuclear weapon, through inaction from the international community, is an option that needs to be considered by world leaders.

²⁴ Steven Emerson & Joel Himelfarb, "Would Iran Provide A Nuclear Weapon to Terrorists?" in *Focus Quarterly* 3 (Winter 2009)

²⁵ Mark McDonald, "North Korea Suggests Libya Should Have Kept Nuclear Program," *New York Times* (March 24, 2011)

IV. Option 2 - Sanctions

Sanctions have been supported on a unilateral and multilateral basis as a means for curbing Iranian access to nuclear materials necessary and dissuading it from developing nuclear weapons. Despite Iran's persistent claims that its nuclear program is strictly for peaceful purposes, its development seems to go beyond what is necessary for mere energy production. The goal of sanctions is to put pressure on Iran's ability to continue its nuclear program. Sanctions are a method of "hard" power and they operate within the framework of international liberalism, which emphasizes diplomacy as the solution to international problems. While sanctions do have an effect in delaying nuclear development, thus far, they have not been very effective in suppressing Iran's ambitions for nuclear weapons.

On July 31, 2006, the U.N. Security Council passed Resolution 1696, which "demanded that Iran suspend all enrichment-related and reprocessing activities," including research and development, and gave it one month to do so or face the possibility of economic and diplomatic sanctions.²⁶ Iran refused to comply with this resolution. Facing Iranian defiance, the Security Council unanimously passed Resolution 1737 on December 23, 2006, which imposed "sanctions on Iran's trade in sensitive nuclear materials and technology."²⁷ This tougher round of sanctions demanded that Iran halt its uranium enrichment program within 60 days, with threats of further sanctions. Again, Iran defied the will of the Security Council, so following the 60-day grace period the Security Council passed Resolution 1747 on March 24, 2007. This intensified the previous sanctions, named specific officials as the targets of the sanctions, and added additional sanctions against Iranian financial institutions.²⁸ On March 3, 2008, the U.N. passed Resolution 1803, which reaffirmed the previous sanctions and asserted greater IAEA rights to verify information provided to it by Iran.²⁹ With Iran continuing its violations

²⁶ Security Council 5500th Meeting (AM), "Security Council Demands Iran Suspend Uranium Enrichment by 31 August, Or Face Possible Economic, Diplomatic Sanctions" (July 31, 2006)

²⁷ Global Policy Forum, "UN Sanctions Against Iran"

²⁸ Ibid.

²⁹ Security Council 5848th Meeting (PM), "Security Council Tightens Restrictions on Iran's Proliferation-Sensitive Nuclear Activities, Increases Vigilance Over

of international calls for an end to its nuclear program, the U.N. passed Resolution 1929 on June 9, 2010. This resolution imposed additional sanctions, expanded an arms embargo, and tightened restrictions on “financial and shipping enterprises related to ‘proliferation-sensitive activities.’”³⁰

These sanctions have seemingly succeeded in complicating Iran’s ability to further develop its nuclear program, but have failed to change Iran’s behavior. Iran continues to assert its right to enrich uranium without external interference. In response to international criticisms, Iran has argued that “the countries that have backed sanctions have provided no evidence” to prove that it hopes to construct nuclear weapons and it points to a 2007 U.S. National Intelligence Estimate that declared that Iran had ended its nuclear weapons research in 2003.³¹ The fact of the matter is that although its excessively large amounts of uranium enriched is suspicious, that alone is not sufficient proof that Iran is aiming towards nuclear arms. Iran has continued to demonstrate its unwillingness to allow foreign states to determine what it claims is a purely domestic issue. With a great deal of assistance from North Korea, Iran is not dependent on the West for nuclear assistance. Thus, sanctions that limit Iranian trade of nuclear materials and technologies with participating U.N. states have had little effect on Iran’s program.

The failure of sanctions is evident by the September 2009 revelations about the construction of a uranium enrichment facility buried inside a mountain near the city of Qom.³² This facility, which can hold approximately 3,000 centrifuges, is disturbing because it appears to be directly aimed at developing nuclear weapons. Iran has openly admitted that the site is too small for peaceful civilian purposes, leading the international community to believe that it will be used for arms development. Additionally, the facility is “located in an underground tunnel complex on the grounds of an Islamic Revolutionary Guard Corps Base.”³³

Iranian Banks, Has States Inspect Cargo” (March 3, 2008)

³⁰ Security Council 6335th Meeting (AM), “Security Council Imposes Additional Sanctions On Iran, Voting 12 In Favour To 2 Against, With 1 Abstention” (June 9, 2010)

³¹ Global Policy Forum, “UN Sanctions Against Iran”

³² Ibid.

³³ Nima Gerami & James M. Acton, “What Else Is Iran Hiding?” *Foreign Policy* (September 28, 2009)

The close proximity between the nuclear facility and Iran’s military institution gives further credence to the belief that it is being used for military and weapons purposes. Sanctions failed to stop Iranian construction of new nuclear facilities, and seem incapable of altering its hostile behavior.

Following the exposure of the Qom facility and with increasing fears of more undisclosed nuclear facilities, the international community felt greater urgency in developing stricter sanctions to limit Iran’s access to nuclear materials and technologies. The U.S. has sought to build a stronger coalition in the U.N. to strengthen future sanctions. In the past, China and Russia, two veto-holding members of the Security Council, have resisted calls for tough, multilateral sanctions. China, an increasingly self-confident power on the world stage uses “its diplomatic weight to protect countries that Beijing considers to be its allies.”³⁴ It considers Iran an ally out of economic convenience because it has sought increased trade with Iran, particularly in the sale of oil, which is in great demand because of China’s rapid development. Russia has also argued against the use of sanctions on the grounds that sanctions are counterproductive because they force Iran to be more secretive. Instead, it believes that “all efforts must be focused on supporting the negotiating process.”³⁵

Following the discovery of Qom, China and Russia began cooperating with the West on the Iran issue. With their support, the Security Council was able to pass Resolution 1929, which was much stricter than previous sanctions by expanding sanctions with hopes of applying pressure on the Iranian economy. In order for more successful sanctions, there must be greater support from the rotating members of the Security Council who have traditionally called for softer methods of engagement with Iran. Members of the 2010 Security Council, such as Brazil, Turkey, and Lebanon advocated for the use of diplomacy rather than sanctions. They argued “the sanctions that have been slapped on Iran have not made the Iranian government more responsive to the demands of the Security Council and the IAEA. However, these sanctions have caused Iranian

³⁴ Louis Charbonneau, “Analysis – More Iran steps seen tougher sell on new UN council,” *Reuters Africa* (December 31, 2010)

³⁵ Megan K. Stack, “Russia opposes tougher sanctions against Iran,” *Los Angeles Times* (October 14, 2009)

civilians much hardship.”³⁶ Members of the 2011 Security Council, such as India and South Africa, have also demonstrated reluctance towards increasing new rounds of sanctions on Iran, which would be targeted the Iranian energy sector.³⁷ They oppose sanctions on the grounds that they would severely disrupt the international oil markets. For developing economies, maintaining Iran’s oil output is of greater importance than addressing the threat of nuclear weapons. Nuclear proliferation is not as serious of a concern for the developing world as it is for the developed world. Developing countries are more concerned with economic development rather than insecurity from the nuclear powers. Iran’s oil production is its trump card against calls for stricter economic sanctions. The countries that oppose further sanctions against Iran question the legitimacy and efficacy of the sanctions. While the developing countries on the Security Council have expressed hesitation in applying sanctions, President Obama’s openness to engagement with Iran will help to ensure that India, South Africa, China, and Russia keep an open mind in regards to further U.N. sanctions.³⁸

The consensus amongst the U.S. and its allies is that multilateral sanctions have slowed Iran’s nuclear program and its suspected effort to develop nuclear weapons. Officials in the United States, Europe and Asia believe that their international campaign “has restricted Iran’s ability to procure the raw materials needed to build an atomic bomb,” particularly “carbon fiber and a particular high-strength steel.”³⁹ The inability to acquire these materials has stalled Tehran’s efforts of constructing advanced centrifuge machines that could accelerate the enrichment of uranium for nuclear weapons. While sanctions have slowed Iran’s uranium enrichment, it is clear that by no means have they achieved their goal of subduing Iran’s ability to construct nuclear arms. Estimates from the U.S. and Israel show that Iran “already has enough low-enriched uranium stockpiled to create as many as four atomic weapons.”⁴⁰ There is also

the worry that Iran could possibly have advanced centrifuges at nuclear facilities that are unknown to the IAEA. Israel still suspects that Iran will be able to build a nuclear weapon within the coming few years. Sanctions are unable extinguish the ambitions for nuclear weapons, which is largely based on security concerns and hegemonic desires. Instead, all it can do is slow Iran’s process by limiting its access to the nuclear materials market. If Iran wants nuclear weapons, sanctions alone will not be able to prevent it from doing so.

V. Option 3 - Covert Operations

In addition to sanctions, covert operations have been used to slow Iran’s nuclear program by sabotaging nuclear equipment and diminishing expertise. Covert operations against Iran’s nuclear facilities have utilized cyber-attacks against nuclear facilities, and attacks against Iranian nuclear scientists. Given the nature of covert operations, it is unknown what specific actions have been taken and their effectiveness. Known covert operations against Iran’s nuclear program are the Stuxnet computer worm, and the Mossad’s assassinations of Iranian scientists.

Stuxnet, known as the world’s first cyber missile, is a computer worm that is designed to destroy the power supplies used in nuclear fuel-refining centrifuge systems. It is a very mysterious computer virus because of its encrypted nature and unknown origins. A Belarus antivirus company discovered it in June 2010 and its purpose of destroying specific nuclear targets was not understood until September of that year. While the origins of the worm are still unknown, many computer scientists, nuclear experts, and former intelligence officials believe that Stuxnet is a joint project between the Americans and the Israelis, possibly with some help, from the Germans and the British.⁴¹ In the last year of his presidency, President George W. Bush “authorized new covert action intended to sabotage Iran’s suspected effort to develop nuclear weapons.”⁴² This authorized covert action is believed to be the origins of the development of Stuxnet.

³⁶ Global Policy Forum, “UN Sanctions Against Iran”

³⁷ Charbonneau, “Analysis – More Iran steps seen tougher sell on new UN council”

³⁸ Ibid.

³⁹ Jay Solomon & Charles Levinson, “Sanctions Slow Iran’s Nuclear Capability,” *Wall Street Journal* (January 8, 2011)

⁴⁰ Solomon & Levinson, “Sanctions Slow Iran’s Warhead Capability”

⁴¹ William J. Broad, John Markoff & David E. Sanger, “Israeli Test on Worm Called Crucial in Iran Nuclear Delay,” *New York Times* (January 15, 2011)

⁴² David E. Sanger, “U.S. Rejected Aid for Israeli Raid on Iranian Nuclear Site,” *New York Times* (January 10, 2009)

Following its complex trail, researchers in California and Germany have discovered how Stuxnet operates. They believe that the worm gains access to computer systems using an array of devious exploits. Upon entry, it “searches for and infects only a specific Siemens-made programmable logic controller (PLC) performing specific functions.” Upon finding these targets, the worm hunts for “identification numbers unique to a special kind of ‘frequency converter drive’ made by only two firms in the world: one headquartered in Finland, the other in Tehran.” Of these frequency converter drives, Stuxnet only targets those operating at speeds resembling those that centrifuges must achieve in order to separate and concentrate uranium to produce nuclear fuel. The worm can either bring these centrifuges to “a grinding slowdown or an explosive surge” by forcing it to constantly swing its speeds from extremes for a long period, which causes the centrifuge to break down.⁴³

In 2009, Stuxnet hit Iran’s uranium enrichment facility at Natanz, home to about 5,000 centrifuges. The worm damaged and forced the replacement of about 1,000 centrifuges. Through IAEA cameras installed at the plant, U.N. inspectors recorded “workers hauling away crate after crate of broken equipment.”⁴⁴ Initially, analysts who studied the Stuxnet attack on Natanz estimated that the attack set Iran’s nuclear program back two years. They claimed, “this was nearly as effective as a military strike, but even better since there are no fatalities and no full-blown war. From a military perspective, this was a huge success.”⁴⁵ The worm was able to destroy Iranian machines and its unknown origins prevent the Iranian regime from retaliating.

Despite the initial enthusiasm about the efficacy of Stuxnet in slowing Iran’s nuclear program, it appears that the worm was less effective against the nuclear program than believed. Following the attack, Iranian scientists engaged in a “feverish – and apparently successful – effort” to contain the damage done by Stuxnet and to replace

the damaged parts despite sanctions against Iran’s purchase of equipment used in uranium enrichment. According to an anonymous Western diplomat with access to confidential IAEA reports, Iran seems to be have been able to “maintain a constant, stable output” of low-enriched uranium despite the damage to the centrifuges. In fact, a February 2011 IAEA report shows that Iran maintained “steady or even slightly elevated production rates” at the Natanz enrichment plant in 2010. The damage to Iranian equipment had little to no lasting effect on the rate of Iran’s uranium enrichment.⁴⁶

Although Iran seemed to have been able to easily and quickly replace the damaged centrifuges, Stuxnet was not a complete failure. Iran is believed to have only “finite supplies of certain kinds of high-tech metals needed to make the machines” due to the trade restrictions from the U.N. sanctions.⁴⁷ The forced replacement of the Natanz centrifuges strains Iran’s limited amount of nuclear supplies, and also diminishes its ability to construct new machines. Furthermore, the specificity of the worm’s encoding seems to imply that its designers had intimate knowledge of the Iran’s centrifuges. Despite Iran’s attempted veil of secrecy surrounding its nuclear program, Stuxnet demonstrates the level of foreign intelligence about Iran’s nuclear program. International computer analysts who have examined the worm believe that it was created with access to inside information. Moreover, the computers at Natanz that control the centrifuges are not connected to the Internet, which means that Stuxnet could only have been introduced through a hand-held USB device.⁴⁸ Natanz, and possibly other nuclear facilities, must have been infiltrated by foreign intelligence agencies. Iran’s cloak of secrecy has failed to protect its nuclear program from foreign interference.

In addition to the Stuxnet cyber-attacks, the Mossad, Israel’s intelligence agency, is believed to have implemented assassination plots against key Iranian nuclear scientists. On November 29, 2010, Majid Shahriari, Iran’s top nuclear scientist and senior manager of Iran’s nuclear effort, was killed when “an explosive charge placed in his car was detonated by remote control after he climbed into

⁴³ Mark Clayton, “How Stuxnet cyber weapon targeted Iran nuclear plant,” *Christian Science Monitor* (November 16, 2010)

⁴⁴ Joby Warrick, “Iran’s Natanz nuclear facility recovered quickly from Stuxnet cyberattack,” *Washington Post* (February 16, 2011)

⁴⁵ Yaakov Katz, “Stuxnet virus set back Iran’s program by 2 years,” *Jerusalem Post* (December 15, 2010)

⁴⁶ Warrick, “Iran’s Natanz nuclear facility recovered quickly from Stuxnet cyberattack”

⁴⁷ Ibid.

⁴⁸ Roula Khalaf & James Blitz, “The Sabotaging of Iran,” *Financial Times* (February 11, 2011)

the vehicle.”⁴⁹ The details of this assassination are characteristic of the Mossad, which has carried out similar plots in the past. The effects of Shahriari’s assassination have been twofold: it caused the immediate loss of “operational expertise and detailed knowledge of an effort that has gone on for decades” and has also served as a threat to other scientists working on Iran’s nuclear program.⁵⁰ In addition to Shahriari’s assassination, Ali Mohammadi, an Iranian quantum physics expert, was killed when a bomb exploded near his house. Many suspect that these assassinations were committed by the Mossad. Both Shahriari and Mohammadi were Iranian participants to the Sesame project, a scientific study of atomic structures that brings together scientists from Israel, Arab states, and Iran.⁵¹ It is likely that is through the Sesame project that the Israelis came into contact with the key Iranian nuclear scientists. The Mossad’s assassinations of nuclear scientists have diminished Iran’s intellectual supply of scientists needed for a successful nuclear program.

In conjunction with the Mossad’s attacks against Iranian nuclear scientists, the United States has worked covertly to prevent Iranian construction of nuclear weapons. In 2007, the U.S. escalated covert operations in Iran through the C.I.A. and the Joint Special Operations Command. These operations were primarily designed to destabilize the Iranian regime through support of dissident organizations and to gather undermine Iran’s nuclear program. While the exact nature of U.S. involvement in Iran is unknown to the public, it correlated to an increase in violence in Iran.⁵²

Publicly known covert operations have been successful in limiting Iran’s access to nuclear technology and operational knowledge. Because of its clandestine nature, covert operations are advantageous because they do not force Iranian retaliation. They are effective tools for disrupting Iran’s nuclear progress beyond the public eye, decreasing the pressure on Iranian leaders to retaliate. Yoel Guzansky, former head of the Iran desk on the Israeli National Security Council, praises covert operations as the moderate option between sanctions and a use of military force

⁴⁹ Time, “Is Israeli’s Mossad Targeting Iran’s Nuclear Scientists?” (November 30, 2010)

⁵⁰ Time, “Is Israeli’s Mossad Targeting Iran’s Nuclear Scientists?”

⁵¹ Khalafa & Blitz, “The Sabotaging of Iran”

⁵² Seymour M. Hersch, “Preparing the Battlefield,” *The New Yorker* (July 7, 2008)

because it delays Iran’s actions and creates time for sanctions and diplomacy to work.⁵³ Covert operations have successfully slowed aspects of Iran’s nuclear development, but have not yet completely stopped it. They may be incapable of changing Iranian motives for much of the same reasons sanctions can’t. They do not address Iran’s desires for nuclear weapons, but instead merely places hurdles between Iran’s nuclear program and the capability to construct weapons.

VI. Option 4 - Preventive strike

As opposed to the safer options of sanctions and covert operations, preventive strike is a very controversial and risky option proposed for dealing with Iran’s nuclear program. The most likely form of a preventive strike would be an air raid against key Iranian nuclear sites to destroy enough centrifuges and other nuclear machines. The most likely candidates to conduct such attacks are Israel and the United States. Israel, when threatened, has a history of preventively attacking facilities in the Middle East believed to be a part of a nuclear weapons program. In 1981, Israel struck Iraq’s nuclear facilities in Osirak, and in 2007, attacked Syria’s Deir Ezzor Research Station. The U.S., under both the Bush and Obama administrations, has stated reluctance towards the use force against Iran’s nuclear program, in spite of encouragement to do so by some Arab allies. Israel, on the other hand, has the military capability to unilaterally destroy Iranian nuclear facilities and appears to be seriously considering it. Preventive strikes are the quintessential course of action within the realism paradigm. Uncertain of Iranian motives, Israel may use the option of eliminating the threat through use of force.

Israel’s history of preventive strikes against enemy nuclear facilities gives important insight into potential future strikes on Iran. In 1981, Iraq was in the midst of constructing a 70-megawatt uranium-powered reactor at Osirak, a facility that Israeli intelligence believed was designed for the production of nuclear weapons. Acknowledging Iraq’s history of aggression against Israel, Israel decided to use its air force to prevent Iraq from attaining nuclear arms. On June 7, 1981, Israeli F-15 interceptors and F-16 fighter-bombers struck and decimated the Osirak nuclear reactor near Baghdad, about 700 miles from Israel. Justifying

⁵³ Khalafa & Blitz, “The Sabotaging of Iran”

this air strike, the Israeli government stated, “the atomic bombs which that reactor was capable of producing whether from enriched uranium or from plutonium, would be of the Hiroshima size. Thus a mortal danger to the people of Israel progressively arose.”⁵⁴ The potential for such a deadly weapon falling into the hands of Israel’s enemies justified the use of force for self-defense, and this justification would apply to any preventive strikes on Iran. While Israel was able to successfully destroy the Osirak reactor relatively easily, preventive strikes were by no means successful in curtailing Iraq’s nuclear weapons program in the long run.

The strike against Osirak was successful in the immediate destruction of the Osirak nuclear reactor. While important, a reactor is not an entirely essential part of a successful nuclear weapons program. The Iraqis “would have needed to construct a separate plutonium reprocessing plant” if it were interested in immediately attaining nuclear weapons.⁵⁵ There was no evidence of Iraq having a plutonium reprocessing plant or that it had plans to construct one in the immediate future. The strike against the Osirak nuclear reactor was unnecessary because Iraq wasn’t in a position to construct a nuclear weapon. Israel prevented the construction of nuclear weapons that was not yet imminent. In doing so, the attacks revealed Israel’s means of attack and its strong intent in doing so. Rather than suppressing Iraq’s nuclear intentions, the repercussions of the strikes were that they forced Iraq to develop nuclear technology more secretly, hampering Israeli intelligence of future Iraqi nuclear developments.

A decade after the attacks on Osirak, the U.S.-led coalition forces fought Iraq in the first Persian Gulf War. After this conflict, U.N. inspectors “unearthed a huge infrastructure for nuclear development that had been completely unknown to Western intelligence before the war.”⁵⁶ Unbeknownst to the international community, Iraq had secretly been developing a nuclear program that came close to the production of a nuclear weapon. The Israeli strikes against Osirak, while delaying Iraq’s nuclear development a decade earlier, did not deter Iraq from continuing its nuclear program. The fact that Iraq was capable of

almost constructing nuclear weapons demonstrates how preventive strikes failed to stop a nuclear weapons program.

In an echo of the Osirak strike, Israel attacked the Deir Ezzor Research Station in Syria in September 2007. The target of the attack was a site that foreign intelligence analysts believed was a partly constructed nuclear reactor that was “modeled on one North Korea has used to create its stockpile of nuclear weapon fuel.”⁵⁷ It appeared to have been the beginning stages of a Syrian nuclear program with North Korean assistance. A senior Israeli official speaking on the condition of anonymity stated that the strike’s intention was to “re-establish the credibility of our deterrent power.”⁵⁸ The message behind the strike against the Syrian reactor, which did not pose an immediate threat to Israel, seemed to be directed towards Iran. It asserted Israel’s willingness to use preventive force to assure its security.

While Israel has relied on the use of preventive strikes against hostile nuclear-ambitious states in the past, it is much less of an attractive option for the current situation with Iran. Iran poses a uniquely dangerous threat to Israel. It understands Israeli military capabilities and its means of attacks. In response to the potential of an Israeli air strike, Iran’s nuclear program entails a large nuclear complex that is carefully concealed, extensively spread throughout the country, and provides multiple routes to nuclear weapons capability.⁵⁹ The immensity of the nuclear network and the geographic diversity of its various facilities make difficult the prospects of successfully halting the nuclear program through preventive strikes. Unlike Iraq and Syria, Iran’s nuclear program does not provide a clear, singular target because it has diversified the locations of its most critical nuclear fissile material production sites. These critical nuclear fissile material production sites are the uranium conversion facility in Isfahan, the large uranium enrichment facility at Natanz, and the heavy water plant and plutonium production reactors under construction at Arak. These facilities are each about 950 to 1,000 miles away from Israel,

⁵⁴ David E. Sanger & Mark Mazzetti, “Israel Struck Syrian Nuclear Project, Analysts Say,” *New York Times* (October 14, 2007)

⁵⁵ Ibid.

⁵⁶ Whitney Raas & Austin Long, “Osirak Redux? Assessing Israeli Capabilities to Destroy Iranian Nuclear Facilities,” *International Security* 31 (Spring 2007) 7-33

⁵⁴ BBC News, “1981: Israel bombs Baghdad nuclear reactor”

⁵⁵ Richard K. Betts, “The Osirak Fallacy,” *The National Interest* 83 (Spring 2006) 22-25

⁵⁶ Ibid., 22

the far end of the distance margin that Israeli fighter planes can safely reach. These sites are imperative to the success of a nuclear weapons development, and “destruction of these facilities would have the greatest impact on Tehran’s ability to manufacture nuclear weapons.”⁶⁰ In addition to the difficulty of striking the numerous facilities throughout Iran, the types of facilities discourage any use of force.

The uranium conversion facility at Isfahan would be a deadly target for a military strike because of the nature of the site. It is estimated that tons of uranium exist at Isfahan in chemical form, and preventive strikes against this facility “could result in the release of tons of UF6, UF4, and other fluorine and uranium products into the atmosphere,” which would very likely “result in significant production of hydrofluoric acid, a highly corrosive chemical” in the atmosphere.⁶¹ Unleashing this chemical into the atmosphere could prove deadly for the 1.5 million residents of Isfahan, Iran’s third largest city, and can affect the lives of millions of people in the region. The results of an attack on the Isfahan uranium conversion facility would be detrimental to not only Iran’s nuclear weapon ambitions, but also to the lives of millions of innocent people as well.

Like Isfahan, air strikes against Natanz and Arak may not be necessary. Preventive strikes against Natanz would not be dangerous in the same ways as strikes against Isfahan would be. However, preventive attacks may be superfluous given Stuxnet’s infiltration of Natanz. The fact that the worm must have been introduced manually by a USB drive indicates that outside intelligence has access to Natanz. It has been demonstrated that the machines at Natanz can be attacked and destroyed without the controversy of a preventive strike. Arak is currently in the process of constructing a heavy water plant and plutonium production reactors. These heavy water reactors pose a great risk of plutonium proliferation because those produced by the Arak reactors would be weapons grade. It is suspected that the Arak facility will be used to recover plutonium from fuel spent in the process of producing weapons-grade plutonium. A reliable means of recovering nuclear fuel would be vital for the successful construction of nuclear weapons. Its destruction would “significantly slow Iran’s future ability to plutonium” and limit Iran’s

supply of nuclear materials.⁶² However, its construction is not yet complete, and it is possible that those Arak will be infiltrated in the same manner as Natanz. It is too early to contemplate preventive strikes against Arak because it has yet to be determined how large of a threat it actually poses. Preventive strikes against Natanz and Arak are neither currently necessary nor smart given that they can be attacked in other ways. In the immediate future, there isn’t much urgency for the destruction of these nuclear missile production sites.

While Israel’s air force has demonstrated its capability of successfully destroying nuclear facilities in Iraq and Syria, Iran poses strategic difficulties. Israel needs U.S. cooperation in case of an airstrike because it controls much of the air space between Israel and Iran. This cooperation may be difficult given the unpopularity of preventive strikes. Iran also has greater means of disrupting Israeli attacks. Iran’s air defense capabilities consist of three elements: aircrafts, surface-to-air missiles, and antiaircraft artillery.⁶³ While these defenses are technologically outdated, with most dating back to before the Islamic Revolution, they cannot be fully discounted. They are still capable of disrupting the flight of Israeli aircrafts. In addition to its outmoded air defense system, the Iranian air force has forty modern MiG-29s jet fighter aircrafts that can engage in air-to-air combat with the Israeli air force. In any event, Israel will not be able to strike Iran’s nuclear facilities easily.

Despite Iran’s technological disadvantages, it has two important advantages against an Israeli airstrike in Iranian airspace. First, Iran’s aircrafts would be operating in airspace near their docking bases. This will allow them to carry more fuel than their Israeli counterparts, giving them an edge during instances of protracted air-to-air combat. Second, Iranian aircrafts can rely on Ground Control Intercept radars to disclose the position of Israeli aircrafts. This would allow the Iranian air force to begin engagement with its Israeli adversaries from a favorable position, such as attacking from behind.⁶⁴ These advantages can create difficulties for Israel if it were to decide to attack Iranian nuclear sites.

⁶⁰ Ibid., 15.

⁶¹ Ibid., 14.

⁶² Ibid., 21.

⁶³ Ibid., 22.

Prior to the pro-democracy protests across the Middle East, several Arab states showed support for preventive strikes against Iran. According to U.S. State Department cables revealed by WikiLeaks, Arab states such as Saudi Arabia, and Bahrain lobbied for the U.S. to strike Iranian nuclear facilities. Referring to Iran's suspected nuclear weapons program, Saudi King Abdullah asked the U.S. to "cut off the head of the snake" before it was too late.⁶⁵ Arab states that have traditionally conflicted with Iran fear the power and influence that nuclear weapons would afford Iran. The cables also highlight "Israel's anxiety to preserve its regional nuclear monopoly" and its readiness to strike Iran alone.⁶⁶ Arab states neighboring Iran have shown support for preventive attacks. Without Arab opposition to preventive strikes against Iran, Israel may find unilateral preventive strikes very appealing.

While they are advantageous in some respects, preventive strikes are too risky to be carried out. Israeli air strikes would likely overcome Iranian defenses and succeed in the destroying Iranian nuclear targets, but doing so will most likely demand retaliation from the Iranians. Given Iran's views on Israel, tensions would rapidly escalate and force Iran to launch counter attacks that can escalate into an all-out war. War with Iran is particularly perilous because it "has the largest and most ideologically committed military of any state in the Persian Gulf region."⁶⁷ In addition to the Iranian military, in case of a war, Hamas and Hezbollah, terrorist groups with close ties to Iran, would likely assist Iran in inflicting catastrophic damage to Israel. War with Iran will also further complicate Israel's prospects of peace with its surrounding countries. These risks are not worth the delaying of an Iranian nuclear program. Preventive strikes are unable to suppress Iran's nuclear ambitions. They do not address the reasons for Iran's nuclear ambitions, and could only slow its nuclear advancement. Like Iraq, Iran can covertly continue its nuclear development after the

⁶⁵ Borzou Daragahi & Paul Richter, "Iran 'must be stopped': Arab leaders implored U.S. to attack, WikiLeaks disclosures show," *Los Angeles Times* (November 29, 2010)

⁶⁶ Max Fisher, "WikiLeaks Shines Light on Israel-Arab-Iran Triangle," *The Atlantic Wire* (November 29, 2010)

⁶⁷ George Friedman, "Bahrain and the Battle Between Iran and Saudi Arabia," *Strafor Global Intelligence* (March 8, 2011)

destruction of its most visible facilities if it chose to do so. A clandestine Iranian nuclear program would be even more dangerous to the international community.

Ultimately, the likelihood of preventive strike against Iran's nuclear facilities must be understood through the perspective of Israel's leaders. Despite the shortcomings of preventive strikes, Israel sees the world within the theory of realism. With uncertainty and constant security threats, Israel may find preventive strikes as the only means of halting the potential for Iranian nuclear weapons. Iran's threat to Israel must be understood in the context of the Holocaust and Israeli existential concerns. Iranian threats of annihilation may justify Israeli use of force. While the rational deterrence theory would theoretically prevent a nuclear war, the Israeli leadership may not believe that it applies in the case of Iran. Former Iranian President Rafsanjani has said "the use of an atomic bomb against Israel would destroy Israel completely while [a nuclear attack] against the Islamic countries would only cause damages." Nuclear war between the two may only assure the destruction of Israel, and not Iran. This increases the pressure on Israel to terminate Iran's nuclear weapons program before it achieves weapons capability. The likelihood of preventive strikes hinges on the Israeli leadership's perception of the threat. Further development of Iran's nuclear program with increased threats by Iranian leaders may force Israeli military action.⁶⁸

VII. Option 5 - Regime Change

Regime change is an option to prevent Iran from attaining nuclear weapons by replacing its nuclear ambitious leaders with ones open to diplomacy. Iran's current nuclear program has largely been designed by Ayatollah Khamenei and galvanized by President Mahmoud Ahmadinejad. This regime zealously defends Iran's right to pursue nuclear technology, limits outside knowledge of its nuclear development, and has encouraged violence in the Middle East. Removing these nuclear zealots from power would thwart Iran's ambitions for nuclear weapons.

Attempts at Iranian regime change can come either internally or externally. External attempts to overthrow the current Iranian regime are unlikely. Israel, the state most threatened by the current

⁶⁸ Jeffrey Goldberg, "The Point of No Return," *The Atlantic* (September 2010)

Iranian regime, has demonstrated, at least publicly, little interest in actively overthrowing Iran's government. Doing so would require the use of military force and inevitably lead to a deadly, protracted war and would disrupt the international oil market, among other things. This would lead to condemnation by many in the international community. Overall, external changes to Iran's regime are not a viable option.

An internal regime change is conceivable, and has become increasingly plausible given the current unrest in the region. Iran has a history of indigenous revolutions leading to dramatic regime changes. In the 1979 Islamic Revolution, the Iranian people rose up against the Shah, and replaced the monarchy with a fundamentalist Islamic republic. If Iran's current regime were to be replaced by one less hostile to the international community, the issue of the proliferation of nuclear weapons could be solved diplomatically. Iran's political climate since 2009 seems to indicate that regime change towards democracy is a possibility.

In the disputed 2009 Iranian Presidential election, Mir Hossein Mousavi, a reform candidate, sought to "wrest the presidency and executive power away from radical hard-liners whose term in office had been marked by economic incompetence, foreign-policy adventurism, and an ideological doctrine that included limits on civil right."⁶⁹ Following President Ahmadinejad's electoral victory, Mousavi and other opposition candidates charged that the elections were rigged. Supporters of opposition leaders organized the Green Movement, with hopes of annulling the presidential election. They began protesting in several major cities. Despite the initial peacefulness of the protests, they quickly turned violent. There were numerous reports of "beatings and murders of some demonstrators on Iran's street" by police forces and the Revolutionary Guards.⁷⁰ Violence against protestors transformed the Green Movement from a movement against a disputed election into a movement against Ayatollah Khamenei and the Iranian form of government. With this development, Ayatollah Khamenei "ordered a crackdown on any challenges to his leadership."⁷¹ Many key opposition leaders understood that Khamenei was prepared to

massacre thousands of people, and were dissuaded from organizing large protests. The increased governmental use of force suppressed the 2009 Green Movement protests.

In 2011, following the success of demonstrations in Tunisia that brought down President Ben Ali and in Egypt against President Mubarak, the Green Movement reinvigorated protested throughout Iran. On February 14, 2011, thousands of protestors gathered in the streets of Tehran and other major Iranian cities to demand greater democracy and governmental accountability. However, unlike the preceding protests in Tunisia and Egypt where the governments were reluctant to use excessive force against demonstrators, the Iranian regime resumed its 2009 policy of forceful action to quell anti-governmental protests. The Revolutionary Guard and other security forces were prepared to react with force, and in some locations where protests were planned, "witnesses reported that police officers and baton-holding mercenaries outnumbered the protested."⁷² The state-sanctioned violence against demonstrators was again successful in subduing the protests, at least for the time being.

In addition to its use of violence, the Iranian government greatly restricted the impact of the protests by cutting access between demonstrators and opposition leaders. The two main opposition leaders, Mir Hussein Moussavi and Mehdi Karroubi, disappeared just prior to the February 14th protests. There were contradictory reports over whether they had been jailed or had been placed under extreme house arrest, completely cut off from the outside world.⁷³ It is widely speculated that the Iranian government detained the two with hopes of containing the protest's democratic fervor. If true, this detention of opposition leaders is a risky course of action for the Iranian regime. Cutting access to opposition leaders can limit the strength and direction of the demonstrators, but it can also galvanize protestors by fueling their anger. Ali Afshari, an exiled former student leader, believed that there is fear in the Iranian government that imprisoning the opposition leaders "would give their restless supporters in the Green Movement a

⁶⁹ Hooman Majd, "Think Again: Iran's Green Movement," *Foreign Policy* (January 6, 2010)

⁷⁰ Ibid.

⁷¹ Robert Tait, "Iran's green revolution refuses to wither and die," *The Guardian* (December 27, 2009)

⁷² Liz Robbins, "Protests in Tehran Are Stifled by Security Forces," *New York Times* (February 20, 2011)

⁷³ Neil MacFarquhar, "Mystery Deepens on Status of Iran Opposition Leaders," *New York Times* (February 28, 2011)

new cause to rally around.”⁷⁴ The regime’s actions for suppressing the protests may backfire and engender greater protests, now and in the future.

In addition to the regime’s use of force against Iranian protestors, Iran has also played a role in the protests in nearby countries. It is deeply involved in some protests and recognizes that its interests are connected to the regional protests. One regime that Iran is aiding is Syria, where pro-democracy protests have turned violent. Iran has provided Syria with the equipment necessary to disperse protestors, block the Internet, and track cellphones. These are the techniques that Iran uses to successfully suppress its own protests. Iran’s interest in buttressing the Syrian regime is due to their shared ideology and Syria’s location, which allows Iran to send weapons to Hezbollah and Hamas. Because of the relationship between Iran and Syria, the fall of the Syrian regime would likely galvanize the protests for regime change in Iranian. To protect itself from collapse, Iran’s regime is protecting the Syrian regime.⁷⁵

While suppressing protests in Iran and Syria, the regime has also fomented protests in Bahrain, a small country with a Shi’a majority and a Sunni regime. The protests have increasingly become a proxy war between Iran and Saudi Arabia. With the collapse of the Sunni monarchy in Bahrain, which is supported by Saudi Arabia, Iran would gain greater influence in the region.⁷⁶ The replacement of a Sunni regime with a Shi’a one would increase Iran’s clout of leadership. This perception of regional leadership may decrease public desire in Iran for regime change, and increase the regime’s willingness to use force to quell protests. Increased regional leadership would encourage nationalism in Iran that would shield the regime from some criticisms. Ultimately, it is impossible to determine whether Iran’s anti-government protests will succeed because events in Iran are tied to events throughout the Middle East. The success of the Iranian protesters partly depends on the success of pro-democracy protests elsewhere in the region.

If the Iranian opposition is successful in overthrowing the current regime, the form and ideology of any subsequent regimes is unclear.

While Iranian opposition leaders have embraced democratic values, there isn’t a guarantee that these values will materialize in a new regime. Even if democratic, a new Iranian regime may continue the current regime’s policies for nuclear development. While a democratic government is likely be more receptive to meaningful negotiations on the nuclear issue with the international community, history shows that democratic governments can defy the international community in pursuit of nuclear weapons. For example, the democratic regimes of the Republic of India, the State of Israel, and the Republic of South Africa have all developed nuclear weapons in the past. There is no guarantee a new Iranian regime would give up ambitions for nuclear weapons.

Iranian opposition leaders have defended Iran’s pursuit of nuclear technology, but have stated that they are open to reforming Iranian nuclear policies. During the 2009 elections, Mir Hussein Moussavi stated that if elected, he would be open to negotiations with the United States if it changed its policy towards Iran. He has said that “the consequences of giving up the country’s nuclear program would be irreparable and that the Iranian people support the nuclear program” because the program is “considered a source of national pride.”⁷⁷ His comments demonstrate an unwillingness to give up nuclear development entirely, but also openness to compromise on the purpose of nuclear technology and on the degree of international regulation. The Nuclear Non-Proliferation Treaty does allow for Iran to develop nuclear technology for peaceful purposes, and on the basis of the treaty, diplomacy can be used to halt Iranian production of nuclear arms. Diplomacy can be used to enforce NPT stipulations. If the current protests in Iran successfully lead to a democratic regime, there is an opening for the U.S. and Israel to work with Iran in order to prevent the construction of nuclear weapons.

VIII. Conclusion

Iran’s nuclear ambitions and the fears of the international community can be understood within the framework of realism. Iran wants nuclear weapons primarily for security purposes and to

⁷⁴ Ibid.

⁷⁵ Ariel Zirulnick, “US officials: Iran helping Syria’s Assad put down protests,” *Christian Science Monitor* (April 14, 2011)

⁷⁶ Friedman, “Bahrain and the Battle Between Iran and Saudi Arabia,”

⁷⁷ Associated Press, “Iran’s Reform Candidate, Mir Hossein Mousavi, Says Open to U.S. Negotiations” (April 6, 2009)

attain regional hegemony. These desires do not necessarily pose dangers to the international community. Nuclear proliferation isn't inherently dangerous, only the use of nuclear weapons is. While the international community has been primarily focused on how to stop its nuclear program, the real concern ought to be to limit Iran's nuclear program to strictly peaceful purposes. To ensure that Iran's nuclear program remains peaceful, the reasons for its nuclear ambitions must be addressed. Otherwise, Iran will continue its nuclear program belligerently defiant of the international community.

The best means of ensuring a peaceful nuclear program would involve a combination of some of the options presented. Sanctions are an important aspect of any package of actions for preventing Iran from developing nuclear weapons, but they cannot be the only option. History shows the ineffectiveness of solely relying on sanctions to change undesirable behaviors. Current sanctions on Iran limit its access to nuclear materials, and also threaten the prospects for Iranian hegemony by detaching it from the international community. Despite these benefits of sanctions, they may be detrimental to the average Iranian citizen. This is a major concern brought up by some members of the Security Council who fear that sanctions punish the wrong entities. It may turn the Iranian public against the international community, which can stimulate Iranian nationalism against the outside world. As such, all unilateral and multilateral sanctions need to be targeted strictly against parts of the nuclear program.

Covert operations should also be continued, but done so more prudently. These operations can effectively slow Iran's nuclear development without forcing Iran to retaliate. Insofar as Iran's nuclear program is a source of national pride, publicly known attacks on the program may lead to public desires for retributions, which can escalate into war. In order to avoid this, covert operations must be done more carefully. A very effective use of covert operation would be to use it to lend support to the Green Movement. With regime change increasingly likely due to the state of affairs in the Middle East, international powers can empower protestors.

The Green Movement is no longer merely a movement against the re-election of President Ahmadinejad, but has taken on an ideological fervor for democracy. The success of the Green Movement will be tied to the success of similar revolutions across the Middle East. Success in

countries like Libya and Syria will encourage protesters in Iran. Covert aid to demonstrators can help protect them from excessive governmental force. The clandestine nature of this aid is very important though because protests need to be indigenous in order to succeed. Regime change that doesn't appear to be purely indigenous would be quickly characterized as Western imperialism and would give the Iranian regime greater justification for its use of force.

Regime change to a democratic regime in Iran is the best outcome for the international community. Sanctions and covert operations should try to encourage regime change in order to directly address Iran's nuclear weapons ambition. A truly democratic Iran may not give up its nuclear program entirely, but it will be open to better relations and diplomacy with the West. This can lead to greater international inspections and regulations of Iran's nuclear program to ensure that it is for peaceful objectives. Beyond the issue of nuclear proliferation, regime change would also be very desirable to diffuse other pressing issues in the Middle East that originate from the Iranian regime.

Ultimately if this fails, allowing Iran to develop its nuclear program is better than the option of preventive strikes Iranian nuclear facilities. The risks associated with Iranian nuclear weapons are insufficient in warranting preventive strikes that will surely lead to a deadly war. Iran's vast military and association with terrorist organizations make it a strong adversary in a war. Preventive strikes are also unable to suppress the ambition for nuclear weapons, but can only delay its production. After a strike, Iran can rebuild and resume a nuclear program more clandestinely. A covert nuclear program would be much more deadly than the current Iranian nuclear program with its limited international oversight. The dangers of Iran's nuclear program are not as great as is commonly perceived. While it is in the best interest of the international community to prevent Iran from developing nuclear weapons through diplomacy and safe attacks on its program, if these options fail, the international community must come to terms with Iranian ambitions for nuclear technology and possibly nuclear weapons.

Bibliography

- Associated Press. (April 6, 2009). Iran's Reform Candidate, Mir Hossein Mousavi, Says Open to U.S. Negotiations. Retrieved from: <http://www.foxnews.com/story/0,2933,51277,00.html>
- BBC News. 1981: Israel bombs Baghdad nuclear reactor. Retrieved from: http://news.bbc.co.uk/onthisday/hi/dates/stories/june/7/newsid_3014000/3014623.stm
- BBC News. (July 9, 2008). Iran sends missile test warning. Retrieved from: <http://news.bbc.co.uk/2/hi/7496765.stm>
- BBC News. (May 6, 2010). US leader Biden says Iran may spark nuclear arms race. Retrieved from: <http://news.bbc.co.uk/2/hi/8664636.stm>
- Betts, Richard K. "The Osirak Fallacy," *The National Interest* 83 (Spring 2006) 22-25
- Bright, Arthur. (December 30, 2010). Iran nuclear program delayed, says Israeli minister. *Christian Science Monitor*. Retrieved from: <http://www.csmonitor.com/World/terrorism-security/2010/1230/Iran-nuclear-program-delayed-says-Israeli-minister>
- Broad, William J., Markoff, John & Sanger, David E. (January 15, 2011). Israeli Test on Worm Called Crucial in Iran Nuclear Delay. *New York Times*. Retrieved from: <http://www.nytimes.com/2011/01/16/world/middleeast/16stuxnet.html>
- Bruno, Greg. (March 10, 2010). Iran's Nuclear Program. *Council on Foreign Relations*. Retrieved from: <http://www.cfr.org/iran/irans-nuclear-program/p16811>
- Clayton, Mark. (November 16, 2010). How Stuxnet cyber weapon targeted Iran nuclear plant. *Christian Science Monitor*. Retrieved from: <http://www.csmonitor.com/USA/2010/1116/How-Stuxnet-cyber-weapon-targeted-Iran-nuclear-plant>
- Charbonneau, Louis. (December 31, 2010). Analysis – More Iran steps seen tougher sell on new UN council. *Reuters Africa*. Retrieved from: <http://af.reuters.com/article/ivoryCoastNews/idaFN3111835420101231?pageNumber=1&virtualBrandChannel=0&sp=true>
- Daragahi, Borzou & Richter, Paul. (November 29, 2010). Iran 'must be stopped': Arab leaders implored U.S. to attack, WikiLeaks disclosures show. *Los Angeles Times*. Retrieved from: <http://articles.latimes.com/2010/nov/29/world/la-fg-wikileaks-iran-20101129>
- Emerson, Steven & Himelfarb, Joel. (Winter 2009) Would Iran Provide A Nuclear Weapon to Terrorists? *inFocus Quarterly Journal*. Retrieved from: <http://www.jewishpolicycenter.org/1532/iran-nuclear-weapon-to-terrorists>
- Fisher, Max. (November 29, 2010). WikiLeaks Shines Light on Israel-Arab-Iran Triangle. *The Atlantic Wire*. Retrieved from: <http://www.theatlanticwire.com/global/2010/11/wikileaks-shines-light-on-israel-arab-iran-triangle/22116/>
- Fitzpatrick, Mark. (February 3, 2011). Iran's Nuclear, Chemical and Biological Capabilities – Executive Summary. *International Institute for Strategic Studies*. Retrieved from: <http://www.iiss.org/publications/strategic-dossiers/irans-nuclear-chemical-and-biological-capabilities/press-statement>
- Friedman, George. (March 8, 2011). Bahrain and the Battle Between Iran and Saudi Arabia. *Stratfor Global Intelligence*. Retrieved from: <http://www.stratfor.com/weekly/20110307-bahrain-and-battle-between-iran-and-saudi-arabia>
- Fuhrmann, Matthew & Sechser, Todd S. (January 12, 2011). Would a nuclear-armed Iran really be so dangerous? *Christian Science Monitor*. Retrieved from: <http://www.csmonitor.com/Commentary/Opinion/2011/0112/Would-a-nuclear-armed-Iran-really-be-so-dangerous>

Gerami, Nima & Acton, James M. (September 28, 2009). What Else Is Iran Hiding? *Foreign Policy*. Retrieved from:
http://www.foreignpolicy.com/articles/2009/09/28/what_else_is_iran_hiding

Global Security Newswire. (December 23, 2010). Iran Seeks Nuclear Scientists, Former Diplomat Asserts. Retrieved from:
http://www.globalsecuritynewswire.org/gsn/news_20101223_7294.php

Global Policy Forum. UN Sanctions Against Iran. Retrieved from:
<http://www.globalpolicy.org/security-council/index-of-countries-on-the-security-council-agenda/iran.html>

Goldberg, Jeffrey. (September 2010) The Point of No Return. *The Atlantic*. Retrieved from:
<http://www.theatlantic.com/magazine/print/2010/09/the-point-of-no-return/8186/>

Hersch, Seymour M. (July 7, 2008). Preparing the Battlefield. *The New Yorker*. Retrieved from:
http://www.newyorker.com/reporting/2008/07/07/080707fa_fact_hersh

Katz, Yaakov. (December 15, 2010). Stuxnet virus set back Iran's nuclear program by 2 years. *Jerusalem Post*. Retrieved from:
<http://www.jpost.com/IranianThreat/News/Article.aspx?id=199475>

Khalaf, Roula & Blitz, James. (February 11, 2011). The Sabotaging of Iran. *Financial Times*. Retrieved from:
<http://www.ft.com/cms/s/2/7d8ce4c2-34b5-11e0-9ebc-00144feabdc0.html#axzz1DwZJvbDE>

Lettow, Paul. (April 2010) Strengthening the Nuclear Nonproliferation Regime *Council on Foreign Relations*, 54

MacFarquhar, Neil. (February 28, 2011). Mystery Deepens on Status of Iran Opposition Leaders. *New York Times*. Retrieved from:
<http://www.nytimes.com/2011/03/01/world/middleeast/01iran.html?ref=Iran>

Majd, Hooman. (January 6, 2010). Think Again: Iran's Green Movement. *Foreign Policy*. Retrieved from:
http://www.foreignpolicy.com/articles/2010/01/06/think_again_irans_green_movement?page=full

McDonald, Mark. (March 24, 2011). North Korea Suggests Libya Should Have Kept Nuclear Program. *New York Times*. Retrieved from:
<http://www.nytimes.com/2011/03/25/world/asia/25korea.html>

Milani, Abbas. (December 29, 2010). The Shah's Atomic Dreams. *Foreign Policy*. Retrieved from:
http://www.foreignpolicy.com/articles/2010/12/29/the_shahs_atomic_dreams?page=0,1

Montero, David. (May 4, 2007). Report: Pakistani scientist A.Q. Khan aided Iran. *Christian Science Monitor*. Retrieved from:
<http://www.csmonitor.com/2007/0504/p07s01-wosc.html>

Peterson, Scott. (December 31, 2010). New Year priorities: Tehran focused on turmoil at home, not nuclear program. *Christian Science Monitor*. Retrieved from:
<http://www.csmonitor.com/World/Middle-East/2010/1231/New-Year-priorities-Tehran-focused-on-turmoil-at-home-not-nuclear-program>

Raas, Whitney & Long, Austin. Osirak Redux? Assessing Israeli Capabilities to Destroy Iranian Nuclear Facilities. *International Security* 31 (Spring 2007) 7-33

Robbins, Liz. (February 20, 2011). Protests in Tehran Are Stifled by Security Forces. *New York Times*. Retrieved from:
<http://www.nytimes.com/2011/02/21/world/middleeast/21iran.html?src=mv>

Sanger, David E. & Mazzetti, Mark. (October 14, 2007) Israel Struck Syrian Nuclear Project, Analysts Say. *New York Times*. Retrieved from:
<http://www.nytimes.com/2007/10/14/washington/14weapons.html>

Sanger, David E. (January 10, 2009). U.S. Rejected Aid for Israeli Raid on Iranian Nuclear Site. *New York Times*. Retrieved from: <http://www.nytimes.com/2009/01/11/washington/11iran.html?scp=1&sq=January%202009%20sanger%20bush%20natanz&st=cse>

Security Council 5500th Meeting (AM). (July 31, 2006). Security Council Demands Iran Suspend Uranium Enrichment by 31 August, Or Face Possible Economic, Diplomatic Sanctions. Retrieved from: <http://www.un.org/News/Press/docs/2006/sc8792.doc.htm>

Security Council 5848th Meeting (PM). (March 3, 2008). Security Council Tightens Restrictions on Iran's Proliferation-Sensitive Nuclear Activities, Increases Vigilance Over Iranian Banks, Has States Inspect Cargo. Retrieved from: <http://www.un.org/News/Press/docs/2008/sc9268.doc.htm>

Security Council 6335th Meeting (AM). (June 9, 2010). Security Council Imposes Additional Sanctions On Iran, Voting 12 In Favour To 2 Against, With 1 Abstention. Retrieved from: <http://www.un.org/News/Press/docs/2010/sc948.doc.htm>

Stack, Megan K. (October 14, 2009). Russia opposes tougher sanctions against Iran. *Los Angeles Times*. Retrieved from: <http://articles.latimes.com/2009/oct/14/world/f-g-russia-sanctions14>

Solomon, Jay & Levinson, Charles. (January 8, 2011) Sanctions Slow Iran's Nuclear Capability. *Wall Street Journal*. Retrieved from: <http://online.wsj.com/article/SB10001424052748704739504576067911497955494.html>

Tait, Robert. (December 27, 2009). Iran's green revolution refuses to wither and die. *The Guardian*. Retrieved from: <http://www.guardian.co.uk/world/2009/dec/27/iran-tehran-ayotollah-khamenei-protests>

Time. (November 30, 2010). Is Israeli's Mossad Targeting Iran's Nuclear Scientists? Retrieved from: <http://www.time.com/time/world/article/0,8599,2033725,00.html>

Walt, Stephen M. (October 1, 2009). Iran, arms races, and war. *Foreign Policy*. Retrieved from: http://walt.foreignpolicy.com/posts/2009/10/01/iran_arms_races_and_war

Warrick, Joby. (February 16, 2011). Iran's Natanz nuclear facility recovered quickly from Stuxnet cyberattack. *Washington Post*. Retrieved from: http://www.washingtonpost.com/wp-dyn/content/article/2011/02/15/AR2011021505395_pf.html

Yoong, Sean. (August 3, 2006) Ahmadinejad: Destroy Israel, End Crisis. *Washington Post*. Retrieved from: <http://www.washingtonpost.com/wp-dyn/content/article/2006/08/03/AR2006080300629.html>

Zirulnick, Ariel. (April 14, 2011). US officials: Iran helping Syria's Assad put down protests. *Christian Science Monitor*. Retrieved from: <http://www.csmonitor.com/World/terrorism-security/2011/0414/US-officials-Iran-helping-Syria-s-Assad-put-down-protests>

Red, Why, and Blue: A Study of Upstate Local Politics

Derek Weng

The American political process is established upon a self-adjusting mechanism that weathers down opposing factions, and streamlines different values and needs to a cohesive solution that best addresses the public good. The premise of a democratic government is that elected leaders represent the will of the people. A democratic political system cannot exist without the presence of political parties and their competition to best represent the interests of their constituency. In the ever dynamic state of New York, where an exceptional diversity of geographic, economic, social, and political interests exist in polarizing camps and distinct patterns, it is the role of the state government to ensure that its policies are relatively fair and equal to the differences in needs.

Elections matter in a democracy. At the most basic level, New Yorkers influence state policy by choosing their elected leaders. As New Yorkers participate at the polls every election season to vote for the candidates that they most wish to see in the capital halls of Albany, Washington DC, and local offices, they are placing their sacred trust in a system that promises to uphold the central tenets of democracy- popular sovereignty and majority rule. In a healthy democracy, elected candidates work to represent and resolve constituency concerns. Individual leaders who serve in public office respond to different constituencies, have different priorities, and pursue them with varying levels of dedication to the broader welfare. Political parties back candidates in order to control the state government and institute the party's policies. However, New York's nefarious political balancing act between the two major political parties is a prime example of a declining democracy mired in petty party politics that ultimately harms the people.

The root of all political drama in New York is a conflicting tale of "two states." Like the government, New York is ripped by the historical contentions between New York City and the state's vast entire regions due north. Longstanding economic, social and political differences between Upstate and Downstate New York are among the most vital elements of the state's political and governmental tensions. The common perception is that Downstate, mainly New York City, has

substantially different needs than the rest of the state. The salient Upstate-Downstate divide and the historical mistrust between the two regions have typified much political contentions in the Empire State. Political tradeoffs and compromises are frequently made as legislators strive to appeal and appease the multiplicity of needs between the parties.

Political parties are the principal driving force in instituting public policies. Via a coalition of elected party members, political parties espouse an expressed ideology and a grand plan to materialize their goals through public outreach and in legislation. Political parties are in turn popularized by a base of citizens with shared ideologies who possess the votes needed to elect party candidates to office. In exchange for their vote of confidence, the party will help to represent their voice on the floors of legislatures and political offices. In lieu of this model, perhaps it is prudent to assume that an individual supports his or her political party by a composition of factors that are neither random nor arbitrary. Whether their electoral decision is based on fierce party loyalty, personal political ideologies, or strictly by the merit of the candidates, an individual chooses a candidate and the representing party with faith that the constituent interests will be best represented. In theory, this is an ideal system that democratically engages everyone in its political process through representation of elected officials. But in hindsight, the perfect marriage between political parties and the voters is marred by accusations of mistrust, betrayal, and uncertainty about the future.

With every election season, the downstate heavily hedges its votes on Democratic candidates with the faith that by party tradition they are more sympathetic to their needs for social welfare programs and low-income services. New York City has an overwhelmingly higher percentage of minorities and low income individuals while the upstate region tends to be more homogeneous in ethnic makeup and in affluence. The ethnic minorities tend, on average, to be more liberal than whites and are more concerned with social programs and jobs. Whites, suburbanites, rural groups, and the high income voters align closer with the Republican Party. Likewise, the upstate

populace perennially sides with the Republican Party on its long heralded promises of tax cuts and fiscal conservatism that appeal particularly to their demographic interests.

However, poor economic conditions combined with poor publicity arising from Albany have disillusioned the people of New York. Public opinion on the government and its performance is strongly influenced by the state of the economy. A poll conducted by Marist College Institute for Public Opinion in 2005 found more than half of New York City and downstate suburban residents saying that the state was on the “right track.” But less than a third of Upstate residents agreed, with 62 percent seeing New York on the “wrong track” (Ward, 18). As it turns out, the voters’ perception of the direction the state is heading is greatly shaped by the differences in regional economies.

The great tragedy of the “two state” division is that while the downstate region is protected by the powerful engines of New York City’s world-class economic infrastructure, its upstate sibling is mired in decades of endemic stagnation and decline. While many plausible explanations exist for the iconic decline of the upstate region, the purpose of this paper is a study of the political perspectives of the cause. The lack of a democratic two party system and the insurmountable predominance of the Republican Party, as well as a history of failed reform policies combine to form a formidable contributing challenge to upstate vitality. The upstate electoral base is also committing itself to a precarious disservice by voting for the same party that has failed to revive their economic livelihood throughout recent history.

This analysis of state and local politics will be aided by my personal insight into a state senate race in the 59th district. In more ways than one, this district is a microcosm and a fair representation of the broader regions of Upstate New York that I wish to examine. The 59th State Senate district unfailingly sends Republican officials to the state capital year after year, despite uncertainty as to whether the party is capable of actually reviving the region’s stale economy. The district’s constituents, by the action of their votes, are clearly in favor of maintaining the status quo as opposed to electing a different party that perhaps might be able to bring much needed changes.

The 59th Senate District

My primary interest in local Upstate politics began shortly after joining a local campaign for

New York State Senate of the 59th district in the 2010 general election. While studying political science at the State University of New York at Geneseo, this general feeling of detachment from local politics has prompted me to become more educated and involved in the election scene. While statewide races for governor and US Senators have unfailingly garnered the prominence of press headlines, the majority of local races for state legislatures are for the most part limited to regional media. The student population in large, have ignored the local races all together. The internship has offered me meaningful firsthand knowledge of election politics and local issues important to the district where I currently reside.

The 59th Senate District of New York encompasses all of Wyoming County with select portions of nearby Erie, Livingston (including the college town of Geneseo), and Ontario counties. Currently, Republicans maintain a sizable but shrinking advantage with 77,803 in enrollment over the 63,879 Democrats in the district (New York State Board of Election). In addition, Republican candidates have consistently outperformed Democratic candidates up and down the ticket in recent elections. The retirement of longtime Republican Senator Dale Volker represented a golden opportunity for Democrats to break GOP control of the local senate seat for the first time in thirty plus years.

For two months, I worked as a student intern for Cynthia Appleton, a Democratic hopeful in a formidable Republican district. Her atypical background as a longtime critical care nurse in Rochester’s Strong Memorial Hospital challenged my antiquated notion that most politicians originate as lawyers or businessmen. Appleton’s great political credentials include three terms as the Village of Warsaw Trustee, Chair of Street/Public Works and Fire Department Committees, and a member of the Budget Committee, Warsaw Chamber of Commerce, and the Warsaw Grange 1088. An established creative figure in her community, Appleton is also an award-winning director and an actress in regional theater productions. A refreshing candidate with charismatic qualifications, Cynthia is one who doesn’t follow the usual beaten path of the career politician.

Championing a fiscally responsible platform that aimed to impose a property tax cap and foster small business growth, Appleton was a good example of a Democratic candidate appeasing the moderate swing voters in the district. In fact, the

differences between Appleton and the Republican candidate, Patrick Gallivan, seemed almost negligent. The state budget is excessively bloated and exceedingly inefficient. Naturally, the top concerns of reducing the tax burden on average New Yorkers and cleaning out wasteful or fraudulent state spending are popular bipartisan rhetoric valued by all citizens. Both candidates have included this in their platform. Yet, more cognizant of the voters' frustration with overall bad politics, Cynthia is among the few candidates who were committed to both non-partisan redistricting and campaign finance reform to bring transparency and accountability in election politics.

Since the two platforms converged on many issues, mostly fiscal, the voters theoretically would benefit from either candidate. Both candidates appear to be good representatives of Western New York and are capable of delivering the voices of the Upstate residents to Albany. However, my involvement with the Appleton campaign was entirely motivated by my own partisan bias. Due to the liberal Downstate Democratic doctrines ingrained within me, I was unable to conceive that the Republicans are good for anybody. A proof of my political naivety, I was utterly convinced that the Democrats would better suit the populace of Upstate who are increasingly becoming more dependent on social welfare programs. Upstate, as I know it, is an economically forsaken region clearly in need of new political directions.

Nonetheless, it was an unusually promising campaign for the Democrats in Western New York. Initially, the competition began as a four way race between Cynthia Appleton, Patrick Gallivan, Jim Domagalski on the Independence and Conservative Party lines, and lastly the Carl Paladino-backed Tea Party candidate David DiPietro. The latter three all emerged out of a deeply bitter Republican primary with unwillingness to concede their senate ambitions. The prospect of three Republican candidates each fighting for their share of the voting base essentially guaranteed the seat to the lone Democrat candidate. But on the heels of evaporating support, Domagalski promptly withdrew from the race on October 4, 2010, citing concerns of splitting Republican votes and his fear of a Democrat victory. In a prepared statement for *The Daily News*, Appleton suggested the usual political maneuvering at work, stating "nothing shows better that the Republican side is using its standard bag of tricks than this last minute, back-room maneuvering to get Mr. Domagalski off the ballot by nominating him for a judgeship he admits

he doesn't even want" (Surtel). With the remaining two Republican candidates left in the race, Appleton hedged her bet on voters favoring a change in direction in leadership over the same tiresome Republicans ideals.

In the end however, the same party that has dominated this race for decades prevailed once again. The figures were a landslide for the Republican. When I looked over to the campaign manager on the night of the election, her expression grew tenser by the minute. The local news declared Gallivan the winner long before all the machines have been counted. In the official tabulation, Cynthia Appleton garnered 28.8 percent of the vote while Gallivan won the seat with an overwhelming 55.5 percent (see table 1). The third party candidate, David DiPietro gathered approximately 10 percent of the remaining votes.

The reality was initially too difficult for me to fathom. Seeing the stretch of empty storefronts across the street from the campaign office on Main Street in Warsaw, a feeling of puzzlement and disappointment arose. The defeat of Cynthia Appleton and our Democratic dream has led me to ponder the political mechanisms behind that election outcome. Why do the constituencies of upstate New York consistently vote for the same political party that has failed to realize their campaign promises of job creation and rescuing the economy? In addition, why are the Republicans able to maintain their dominance in the district, as well as other comparable rural upstate districts year after year? These questions have dogged me as I left the abandoned streets of Warsaw that night searching for satisfactory answers.

Upstate's Decline

In order to identify plausible explanations for Upstate's economic and population decline, it is crucial to first comprehend all of its symptoms and ailments. The iconic decline of the Empire State, primarily of the Upstate region, has caused invariable grief to its citizens and politicians alike. As its nickname suggests, New York was once an undisputed forerunner of the American dream. The completion of the Erie Canal in 1825 brought a century of growth and prosperity to both Upstate and Downstate counties. It cut the cost of shipping goods from the manufacturers of Buffalo to New York City by 90 percent, making the latter the preeminent commercial and trading center in the country. At the time, the state's economy was more

than capable of providing jobs for the millions of new residents making home in the Empire State.

Upstate is an unofficial but an exceedingly common vernacular used to differentiate the New York City Metropolitan Area and the entire region due north. Expedient in establishing the biggest urban-rural divide in the state, the term itself objectively capitalizes on the prominence of New York City in relation to the rest of the state. To the Downstate population, there is the city and then there is everything else. Yet, New York is home to numerous other nationally renowned cities like Buffalo, Rochester, and Syracuse. In addition, the political epicenter of the state is located in the capital of Albany, a medium sized city roughly 150 miles north of New York City. So is "Upstate" really a politically correct term that thoroughly encompasses all the economic and political activities outside of Manhattan? Perhaps it is no wonder that Upstate residents are seemingly critical and hostile to the Downstate egotism that has long permeated state politics.

Cornell University economist, Rolf Pendall, argues that the state is made up of essentially three broad regional economies. In his analysis, New York City is the undisputed headquarter of financial activities in the state as well as the nation. The city's economic dynamism affects the entire region's cycle of rise and decline, providing it with much needed growth opportunities as well as great liabilities. The rest of the eastern portion of New York State (which includes Long Island, the northern suburbs, the Hudson Valley, and the Capital District) mirrors rather closely with the prevalent trends of the nation, demonstrating moderate growth in jobs and population. It is the last component of New York, the western and northern frontiers that would essentially rank as the third slowest growing state in the nation if it were to be a standalone state (Schneier, Murtaugh and Pole, 17). Stretching hundreds of miles of uninhabited natural landscape and scattered farms, the region has seen remarkably diminutive economic growth. It is the region that contains the 59th State Senate district and many countless other ones that mirror its vital signs.

The Upstate region as a whole is marked by troubling signs of persistent poverty, evaporation of a young and healthy work force, and less than average wage earning potential. In addition to its economic decline, the region's political clout is shrinking as well. The release of the 2010 Census indicates the loss of two seats in the House of Representatives due to New York's lagging

population growth compared to other states (Barrett and Fleisher). The region that is most likely to lose at least one representative will be none other than Western New York, where population growth has been the smallest. In the 59th State Senate district, Livingston County witnessed a 2.3% decline in population from 2000 to 2009, contrary to the statewide trend of a 3% increase (U.S. Census Bureau). Wyoming and Erie Counties fared even worse, with 4.7% and 4.3% negative population growth respectively during the same period. The district and similar Upstate rural regions are all faced with the daunting prospects of an anemic economy and an aging populace.

The economy of Upstate New York, by nearly all major measures, worsened in the 1990's, and lagged behind both the nation and its own performance in the 1980s. In the fifteen years leading up to 2005, Upstate's economy was among the worst in the country. If considered as a separate state, Upstate's population growth during the 1990s would have ranked 49th in the nation, and just ahead of North Dakota and West Virginia. A study by the Brookings Institution found that nearly 30 percent of Upstate's population increase reflected rise in state prison populations rather than from its natural growth (Pendall, Upstate New York's Population Plateau: The Third-Slowest Growing "State"). Another study released by the Brookings Institution found that Upstate's personal income rose at half the national pace during the 1990s, and most of its new income came from gains in Social Security, public and private pensions and other transfer payments, rather than from new economy growth (Pendall and Christopherson, *Losing Ground: Income and Poverty in Upstate New York, 1980-2000*). Almost all new jobs in Upstate were either on public payrolls, or in sectors such as health care and social services that largely depend on taxpayer funding.

The misfortune of the upstate New York economy is one that has been foretold by the empty storefronts and abandoned factories of countless towns and cities across the Rust Belt states of America. It is the iconic story of the relentless decline of once vital regional manufacturing bases ousted from power. The outsourcing of manufacturing jobs and the structural shift to a service-based economy have depressed real wages and stifled job growth for the past few decades. Much of upstate New York shares the same industrial roots as the Rust Belt states that also border the Great Lakes. It was home to the manufacturing industries of steel,

automobiles, fabricated metal products, and industrial machinery that anchored the economies of former boom towns like Buffalo, Syracuse, Jamestown, Utica, and Schenectady. But these industries were hit hard by the economic recession that started in the early 1970s and the movement of manufacturers and population south to the Sun Belt. A combination of high taxes, labor costs, and energy costs have contributed negatively to the economy, making it almost impossible to foster new growth. As manufacturing jobs were lost and the economy slowly transitioned to service sectors, much of the Upstate failed to recover.

Besides losing more manufacturing capacities, New York did far worse than most other states in keeping and attracting the highest-paying jobs. As employers and residents moved elsewhere, the Upstate cities saw their tax bases stagnate or dwindle. With flat revenues, continual increases in public employee compensation and other expenses, mayors of these cities were forced to impose public service reductions that prompted more residents (especially affluent ones) to depart. Cities such as Buffalo, Syracuse, and Niagara Falls now have “severely constrained revenue streams, high levels of debt and high fixed costs- suggesting that they are so negatively affected by fiscal stress that they have very little local capacity to attain long-term fiscal stability and growth” (Office of the New York State Comptroller, 17). With their population in decline, many Upstate cities saw decline in housing quality and the closure of numerous local businesses. Once-lively neighborhoods have deteriorated into communities of concentrated poverty, often afflicted with high levels of crime and social dysfunction.

In the span of several decades, Upstate New York has downgraded from a stable middle-income region to one with serious income and economic problems. Its declining standard of living compares unfavorably to other parts of the nation (see figure 2). In 1969, per capita personal income in Upstate comfortably exceeded that of the national average. But by 2000, it has trailed the national average by 11 percent (\$26,260 compared to \$29,760) (Pendall and Christopherson, *Losing Ground: Income and Poverty in Upstate New York, 1980-2000*). A prevailing factor contributing to the low wages per capita is unemployment, with only 75% of Upstate men 16 and older working in 1999, compared with 79.5% nationwide. Employment in Upstate New York, including self-employment, rose from just under 3.1 million in 1980 to 3.6 million in 1990, a 17.5 percent increase. An impressive gain by its

own merit, but it was still lower than the national average of 22.1 percent for the same period. By the 1990s, the self-employment gain was less than half as large as it was in the 1980s.

With jobs becoming scarcer, a decent wage earning prospect is also dimming for Upstate (see figure 3). Upstate’s lackluster labor markets present recent college graduates strong wage disincentives in pursuing full time careers in the region. According to the same report from The Brookings Institution, in every age group between 20 and 65 years old, college-educated white men and women who worked fulltime in 2000 earned less than 93 percent of the average national hourly wage. More importantly, the most mobile segment of the population (the white full-time workers with college degrees aged 25-34) earned only 87 percent (males) and 85 percent (females) of the national average wage. Even Upstate’s highest income households failed to earn comparable wages to their national counterparts. The 80th percentile household, whose income is higher than 80 percent but lower than 20 percent of all households, earned about \$74,300 in Upstate in 1999, compared with over \$81,100 nationwide (Pendall and Christopherson, *Losing Ground: Income and Poverty in Upstate New York, 1980-2000*). This 8 percent gap represents a substantial deterioration for Upstate since 1990, when Upstate’s 80th percentile household earned just 3 percent less than its national average.

Like the well-heeled households, Upstate’s poorest households are also in an unsatisfactory position earning lower incomes, on average, than those nationwide. They have become increasingly more dependent on government transfer payments as a source of income. Over half of the increase in Upstate’s per capita income during the 1990s came from growing sources such as social security, government and private-sector pensions, Medicare, and Medicaid. Upstate New York’s traditional status as a low-poverty region no longer holds true. In the 1990s, as poverty declined nationally, it actually crept up in Upstate (see figure 4). In fact, the number of residents living below poverty in Upstate grew by 7.9 percent in the 1990s, a much faster rate of growth than nationwide (6.8 percent) and far outpacing the region’s population growth rate (Pendall and Christopherson, *Losing Ground: Income and Poverty in Upstate New York, 1980-2000*). No longer is the region failing in producing potential new members of the workforce, it is also failing to provide enough living income for its populace.

Upstate's decline is perhaps best demonstrated by the growth of high-poverty tracts, a symbolic representation of population shifts. The number of people below poverty in a tract grows from the influx of poor people and by reduction of established residents' incomes. Upstate's population became more concentrated in high-poverty neighborhoods in the 1990s because 98 tracts with nearly 315,000 residents (more than the entire population of Buffalo city) transitioned from low to high poverty. Upstate's 240 persistent poverty tracts, home to 30 percent of Upstate's poor population in 2000, lost about 69,000 people overall and 13,000 poor residents (Pendall and Christopherson, *Losing Ground: Income and Poverty in Upstate New York, 1980-2000*). This great migration of population out of concentrated poverty has had a negative impact on those left stranded behind in these poor neighborhoods, as they are now surrounded not only by chronic poverty but also by increasing isolation, and in some cases, a loss of threshold densities to support economic, social, religious, and political institutions.

The Upstate-Downstate Divide

Upstate's emblematic decline is accentuated even more by comparison to Downstate's shimmering prosperity. New York City and Upstate New York have viewed each other with suspicion for much of the state's history. Residents in each area tend to believe that their tax dollars subsidize the other. However, detailed studies of the state's revenues and spending reveal that Upstate receives more in state appropriations than it pays in taxes, and that New York City pays more than it receives (Gardner, 2). Nevertheless, the myth has persisted as Upstate continues to resent the city's great influence in regards to state government policy.

The mayor of New York City is a major media figure, capable to command a bully pulpit often exceeding that of the governor (Ward, 524). In 2003, Mayor Bloomberg persuaded state legislators to provide additional emergency budget assistance to the city for employee payroll expenditures. To minimize the negative public outcry, the state legislature opted instead to assume the responsibility for \$170 million in bonds that the city's Municipal Assistance Corporation had originally issued in the wake of the mid-1970s fiscal crisis. The effect was to transfer to taxpayers around the state the burden of repaying debt the city had incurred nearly three decades earlier. An

unfair disservice to Upstate, the already troubled region doesn't need to be burdened even more with downstate's debt. This is just a small speckle in the many powerful ways New York City exercises its command in Albany.

For the residents of Downstate, there are only two categories: New York City, and everything else. The city's political-social establishment looks down its nose at Albany. The attitude of those in New York City toward their state capital (and the rest of the nation) was captured in a famous New Yorker magazine cartoon that showed the world from the elitist perspective of the Upper East Side of Manhattan (see figure 1). In the illustration, there was Manhattan, beyond it a strip of New Jersey and a flat heartland, then the Pacific Ocean in the distance. Albany, as well as the rest of the state, was painfully excluded in the map. Former New York Mayor Ed Koch famously called rural Upstate "a joke", insulting the men's department store suits and the women's "gingham dresses" (Ward, 37). When Democratic Attorney General Eliot Spitzer likened Upstate New York to Appalachia during his 2006 gubernatorial campaign, some Republicans accused him of blatant city-centric elitism. These are just few of the many displays that illustrate Downstate's lack of brotherly attitude towards Upstate.

Failed Policies

It is no surprise that Upstate legislators have historically proposed establishing the region as a separate state, to break away from the perceived negative effects of association with the big city. In a New York Times article, David F. Shaffer, of the Public Policy Institute, largely blamed Upstate's failure to recover on "the dominance exercised by the downstate political culture," an establishment largely synonymous with the Democrats (Roberts). However, political finger pointing is simply incapable of realistically resolving Upstate's economic ailments.

Most of Upstate's candidates for statewide seats have considered it essential to discuss proposed solutions to the region's economic stagnation. But despite detailed promises to reduce taxes and ambiguous projects to reinvigorate the region, the results have yet to be seen. From 1990 to 2005, manufacturing employment declined 20 percent nationally and by 35 percent in Upstate New York. Yet, rust belt states like Illinois, Ohio, and Michigan have enjoyed double-digit percentage gains in overall private-sector employment since

1990, while Upstate's increase was less than 4 percent (Ward, 46). If industry-shift isn't the only factor contributing to Upstate's economic decline, then there must be other factors, perhaps a political one, causing the downturn.

The diminishing figures all point in one direction: the New York State government has largely failed Upstate. Discounting the amount of job losses expected from a structural shift from manufacturing to the service sector, the region has failed to see the new businesses and the economy take form. In a recent study released by the Manhattan Institute, the state of New York's decline is caused by the government itself. "New York's political class spends, taxes, and borrows far more than the national average, consistently ranking at or near the top of the list in every measure of the burden that government imposes on citizens and businesses" (Ward, 46). What New York has is an irresponsible legislature that spends rather than saves for its own sake.

Business leaders and many economists cite high business costs in New York as the key disadvantage in global and national competition for jobs. Extensive anecdotal evidence indicates that business executives and investors react to high costs (Ward, 49). A 2006 report by the Tax Foundation determined New York's state business tax climate as the worst among the fifty states (Gallagher, 5). In a citizen's guide to state finances published by the Office of the State Comptroller, local taxes in New York are 78.7% above the national average, while the state taxes are 8.7% higher (Office of the State Comptroller). New York's debt per capita is over three times the national median. Public school costs, the per capita income tax, and the per capita state-local debt in the state are the highest in the nation, and the state-local tax burden in the Empire State is second only to the burden in Alaska. While business taxes are especially high in the New York City metropolitan region, companies there can usually overcome high costs by virtue of the city's global status in financial and business services, corporate headquarters and other high-value activities. Unfortunately for Upstate region, the businesses there retained the burden of relatively high costs while intrinsically lacking the structural advantages of New York City that retain businesses.

As a result, decades of failed campaign promises from both parties to make Upstate more business friendly and cost competitive have left some New Yorkers skeptical that neither party has the will or the ability to fix the foundering

economy. In the New York Times article, Denis M. Hughes, president of the New York State A.F.L.-C.I.O lamented, "they all had ideas, but none of the ideas worked" (Roberts). In fact, urgent attention needs to be focused on government policies that will effectively address what is essentially a demand deficiency problem at both ends of the income scale. At the high end, Upstate desperately needs new strategies to create more demand for highly skilled workers and offer jobs that offer nationally competitive wages. At the lower end, additional welfare measures are needed to supplement low incomes and to alleviate the growing concentration of poverty. These are bipartisan issues that require bipartisan cooperation in passing the right legislation.

Despite the rhetoric of reform and change, the Governor and the Legislature desperately lack the agency to actually fix New York's chronic ailments. Ironically, Albany is incapable of saving the state because it can't even save itself from mismanagement. Heavily burdened by bureaucracy, fiscal irresponsibility, and bad politics in general, New York is in a substantial amount of debt. The state's budget deficit is at an estimated \$8.2 billion, due in no small part to state spending that has risen by nearly 70%, or \$35 billion, over the past decade (Roberts). The recent financial crisis has exposed the state's overreliance on the downstate economy and on tax revenue from Wall Street. The public's frustration with Albany's inability to govern and the government's preference for New York City has angered many Upstate voters. John Faso, a former leader of Republicans in the State Assembly from Columbia County, blames the deficit on the selfishness of special interests, especially the public employee unions that have crippled the state's economy. "West of Albany and north of Saratoga the state is an economic dead zone," Faso declares in a Wall Street Journal article, "People in downstate have no idea of the anger and frustration against state government in Albany the people in upstate have" (Lind).

Perhaps the best indicators of public discontent with the government can be seen with every election season. The constitution guarantees its citizens the power to elect and reelect politicians that they believe are useful in instituting policies that will benefit their district. Ineffective and unpopular politicians are subjected to being voted out of their office should they fail to live up to their duty. Therefore, given the perpetual dissatisfaction with New York's economic decline, it is my expectation that public discontent will lead to the

systematic defeat of incumbents who have done little to resurrect the local economy. Since a political party is highly indicative of the platform of policies its politician will espouse, it is predicted that party change will result for every district where voters are disillusioned with the previous administration. However during the process of my research, overwhelming evidence suggests otherwise. Despite endemic economic decline, not only are legislative incumbents successful in maintaining their seat, the political parties are also equally successful in maintaining their respective share of power in the state government. Republicans, with a strong Upstate base, have largely retained the majority in the State Senate for 40 years.

This inconsistency can only be explained by the unique political structuring of Albany that has consistently kept the incumbents and the same parties in power. Political parties, aided by gerrymandered districts, have little trouble in securing a majority at the polls which essentially maintains the status quo power structure in Albany. The Republicans have carved a powerful niche in the Senate by utilizing the historical Upstate-Downstate rivalry to persuade Upstate voters that they are the best party to represent their interests. The logic is that a Republican majority in the Senate is the only way to counterbalance the Democrat-controlled Assembly and the recent Democratic governors who heavily tilts in favor of Downstate constituents. Upstate voters are aware of this as they head to the polls in each election season.

A Divided Government

Political parties can either make or break a democratic government. Bipartisan cooperation between the parties can safeguard and expedite the passage of important legislation. Hostile interparty competition, on the other hand, only causes gridlock and accomplishes rarely anything significant or mutually beneficial. Kept to moderation, a competitive party system can function as a democratizing force. Voters benefit when parties compete by offering programs designed for mass appeal and by proving that they can deliver policies through effective governance. In practice however, the two-party system in New York has not worked well in defining real alternatives for the voters or encouraging responsible government policies.

What the Republicans and Democrats have accomplished instead is a heavily divided government and a splintered voter base. The division of voters into the two major parties dates far back into American history. Since the beginning, it was the city man against the rural farmer, the new immigrant against the third-generation American, the Downstate Irish against the Upstate British. What we have today is a state in which patterns of partisanship are clearly evident. Whereas political parties should strive to encompass all interests of the voting population, in reality, they settle to only represent a coalition of people with particular social and political orientations.

The fact that the two parties represent fairly different constituencies within in the state should naturally increase the level of their competition. But despite claims of democracy and fair competition, election outcomes have always been predictable. It has been an electoral tradition for the Democrats to win in New York City (and some urban pockets in Upstate) and for the Republicans to win pretty much everywhere else. In fact, interparty competition for seats in the state legislature became something of a rarity in the 1980s and 1990s. Stonecash showed that since 1900, the average margin of victory (percentage points by which the winner leads the loser) in legislative elections has steadily increased from a little over 20 to the current level of 54 percentage points (Schneier, Murtaugh and Pole, 69). Ironically, the real competition lies within the September primary elections of the respective political parties. For the three-way 59th Senate District primary race, Patrick Gallivan narrowly won with 8,250 votes, with Domagalski coming at second place with 7,289 votes, and lastly DiPietro with an equally impressive 6,591 votes (New York State Board of Election). As the district favors Republican candidates for local races, the official party nomination is highly sought after by the Republican hopefuls.

Both parties have protected and preserved their territory within the legislature. The Democrats have always controlled the State Assembly while the Republicans long dominated the State Senate. The 2006 state legislative elections were relatively typical of the single-party-dominance scenario. In the State Senate, party control changed in only one of sixty-two districts; in thirty four districts, the winning candidate received more than 60 percent of the vote; and ten candidates who ran without major

party opposition. In the State Assembly, only two of the 150 seats changed parties, 116 of those were won by margins exceeding 60 percent, and 42 of these candidates had what amounted to a free ride (Schneier, Murtaugh and Pole, 65). Predictable election results such as these have earned New York the dubious distinction of having some of the least meaningful statewide and local elections in the country.

In addition to uncompetitive local races, New York has one of the lowest rates of voter participation in the nation. In the 2004 presidential election, it ranked forty-eighth in overall turnout (Schneier, Murtaugh and Pole, 75). Few states have a lower turnout in primary elections than New York. A known factor contributing to this phenomenon is the unwillingness of the voters to participate in the races of noncompetitive districts. It is also important to note that common to all ethnic groups, the economically and academically disadvantaged people vote at dramatically lower rates than their wealthier and more educated compatriots. Given the poor economic outlook of the Upstate region, lower income and uneducated voters may have simply given up on voting for a better future. Those who fail to participate in politics are not merely dissatisfied but they are also discouraged to vote. It is possible that the single-party-dominance and the limited choices of candidates have simply made participation in elections meaningless for the public.

Republican Dominance in Upstate

If the political system is failing to offer meaningful and competitive races to voters, then why do the residents of Upstate vote at all? There is a twofold explanation for the Republican dominance of rural and suburban districts in Upstate. In both explanations, the voters are voting entirely on a rational and self-interested basis. For one, the Upstate voters cling to the Republican candidates in retaliation of the further encroachment of Downstate influence in Albany via the Democrat-controlled Assembly. Alternatively, the voters are helplessly manipulated by the political structure in Albany to produce predictable election outcomes. Political parties pour substantial financial and technical resources into key races to influence the voters where the election districts have been gerrymandered, or drawn to incorporate only regions that demographically favor the party. Despite the challenges political gerrymandering may pose to democracy, the

Upstate voters still benefit from having their voice maintained in Albany by the Senate Republicans.

For the majority of the working class voters in Upstate, many of them who might normally be expected to vote Democratic end up voting for Republican candidates in state offices out of hostility to New York City. Compared to its neighboring states in the Northeast, Upstate New York votes more Republican than states with similar demographic profiles. Its Upstate cities are more Republican than cities of equivalent size elsewhere in the country. For many years, Upstate voters have perceived themselves as paying high taxes to pay for New York City social services that were distributed by corrupt or incompetent city politicians (Schneier, Murtaugh and Pole, 69). As of 2006, nine members of the Legislature have been charged with bribery or other forms of crimes within the past five years, all represented districts in New York City where voters are less likely to be aware of their representatives and their activities (Ward 32). This problem exists especially in Downstate districts where the large minority and immigrant population are much less involved in politics.

In the following case study, Upstate politicians seized the opportunity to sharply criticize Albany for playing favoritism with the Downstate politicians. Notably, the Republicans have capitalized on Albany's Downstate bias, by convincing Upstate voters that they are the only party that can help counter New York City's powerful influence.

Case Study: MTA Bailout Benefits Downstate

New York City's immense size alone commands considerable political influence in Albany. Home to more than 8 million residents, about 43% of the state population, the city runs an extensive urban transit system that boasts a daily ridership of approximately 7 million (Metropolitan Transportation Authority). The MTA is a massive state-chartered agency that has constantly been in the news for its lack of funding and threat to terminate services. The agency's budget deficit is a growing crisis for the organization as well as New York City and the state legislature. The MTA currently holds \$31 billion in debt and it also suffers from a \$900 million gap in its operating budget for 2011 (The Drum Major Institute).

The fact is that the governor and the state legislature are directly responsible for the MTA's finances as only they have the power to allocate

funding to the authority. While the MTA has been cutting costs at the margins in terms of reductions in services, its internal efforts have not been enough to close the budget gap and avert an impending fiscal disaster. Much like the myriad state agencies that are over-borrowing and under-performing, MTA's deficit comes at the expense of state taxpayers' dollars. As feared by Upstate residents, Albany has made a contentious decision that unabashedly favors Downstate constituents. The centerpiece of the 2009 MTA bailout plan is a \$1.5 billion payroll tax increase to fund for the agency.

The MTA bailout is a gross negligence of the interest of the Upstate taxpayers. In the process of appeasing the downstate constituents, Albany funds an agency that only services a specific regional populace. In a piece from Senator James L. Seward's blog titled "MTA Bailout Lets Down Upstate", the Upstate Republican (SD-51) raised grave concerns in regards to the bailout for the MTA and the dire consequences his residents. In his argument, most Upstate residents will never set foot on a subway, bus, or train operated by the MTA let alone benefit from it. To further alienate the Upstate citizens, the state vowed to reimburse school districts in the MTA region for their share of the payroll tax to quell the initial public outrage of a tax hike. The reimbursement, costing anywhere from \$60 million to \$100 million, will come from the state's general fund. This means that taxpayers from outside of the MTA region will be helping foot the bill again without exchange for any tangential benefits. He states, "The statewide payroll tax will be devastating for businesses, nonprofits, hospitals, local governments and school districts. Companies outside of New York, thinking about expansion or relocation, will lose interest in moving to New York in short order" (Seward).

There was also a very important omission from the debate over the MTA bailout: Upstate roads and bridges. For decades, the state approved five-year capital plans for the MTA while at the same time passing a proposal to improve roads and bridges in Upstate and on Long Island. Safe, well-maintained highways and bridges are vital to the Upstate economy and daily life. Like the public transportation system that is so vital to the Downstate, Upstate residents depend heavily on highways to get to places. But for the first time in recent years, that parity has been ignored. As the MTA bill authorizes a two year capital plan for mass transit, a road and bridge capital plan for the rest of the state was never discussed.

The exclusion of an infrastructure plan vital to the Upstate is essentially a bad policy that discriminates against the residents of the entire region. Seward states that, "It shouldn't be any real surprise that upstate was left out of this plan. After all, the MTA bailout was put together in secrecy by the same three New York City politicians that crafted the state budget. Both plans were assembled behind closed doors, and both are disastrous to upstate and New York as a whole" (Seward). In the context of the MTA bailout debacle, Republican officials have become an indispensable strong voice for Upstate constituents (despite the fact that they weren't able to change the actual policy). Since the Democrats have a huge constituent base in the city, traditional Republican and swing voters in Upstate are naturally skeptical in voting for a party that might or might not represent their interests. Thus, the Republican Party has become the last savior of Upstate residents as their mouthpiece in lieu of the ever powerful Downstate influence.

Money Influences Elections

The dominance of a party in particular regions can also be explained by a political system that heavily prioritizes winning elections. During the campaign season, the party leaders provide remarkable assistance to individual legislative candidates of their party. In competitive races especially, campaign literature, printing, propaganda, materials on opponents, and so on, are financed and provided by the county party committees. Occasionally, monetary funds are distributed to candidates running in marginal districts. The accomplishments of Republican candidates in winning State Senate seats are the product of successful campaign financing.

Fundraising for campaigns include extensive centralized efforts by the party conferences. The Senate Republican Campaign Committee and the Democratic Assembly Campaign Committee, as the political arms of the majority conferences, can make available hundreds of thousands of dollars to a candidate in an importance race. The campaign committees for the Senate Democrats and Assembly Republicans, the two minority parties, do not raise as much money, but can still provide meaningful funding and technical assistance (Ward, 133). This gives great evidence to the Republican strategy on maintaining the control of the State Senate, their last bastion of power in Albany. The preservation of a Senate majority hinges upon

getting newcomers and incumbents elected. The fact is most rural Upstate districts are under Republican control, and the local county party committees are better organized and financed for the purpose of assisting their candidates. For the party in majority power, it has greater means to assist their candidates by utilizing the research and financial resources of the legislature to these ends.

Judith Hunter, the Chair of the Livingston County Democrats and the campaign manager for Cynthia Appleton, can attest to the power of money in election politics. In a personal interview, Hunter confesses that, “It all comes down to money. Gallivan was able to buy the ads to bombard local airwaves, whereas we ran on a shoestring budget” (Hunter). Indeed, Gallivan’s campaign had a substantial financial advantage. From New York State Board of Election campaign finance reports, Gallivan for Senate raised \$194,453.00 while Cynthia Appleton for Senate raised only \$44,516.26 (New York State Board of Election). Whereas Appleton’s commitment to Campaign Finance Reform has deterred potential special interest money from funding her campaign, Gallivan gained the upper hand of having received much more support from its party committees, PACs, individual donors, and special interest groups. Hunter also lamented the lackluster support from the Erie County Democratic Committee as well as the Ontario County Democratic Committee. Unsure of her chances of victory in lieu of the powerful Republican candidacy, the Senate Democratic Campaign Committees and local Democrat leaders were hesitant in committing money to help her campaign which ultimately hindered her momentum in the race. This is a clear instance where Republican dominance in Senate districts has successfully deterred Democrats from raising and committing the level of money needed to win the race.

Political Gerrymandering & Redistricting

The second part of the analysis focuses on the use of gerrymandering by Republicans to maintain their power in Upstate. Members of the Senate and Assembly in New York have built one of the most effective incumbent-protection systems in the nation. An 1894 constitution was engineered by the Upstate Republicans who controlled the convention to ensure that New York City would never be able to elect a majority of either house of the legislature. This arrangement was accomplished by guaranteeing sixty of the state’s sixty-two counties

at least one seat in the assembly and providing that “no two counties divided by a river” could ever have half the Senate seats (Schneier, Murtaugh and Pole, 79). A constitutional amendment in the 1960s and notable court cases have challenged the plan, mandating a one person –one vote basis in drawing district lines that are strictly equal in population. However, it has not improved the situation of gerrymandering in New York with neither party willing to compromise their greatest shot at winning political offices.

Since the early 1980s, a key element has been the once-every-ten year redrawing of legislative districts that favor majority party incumbents, with the Republicans in the Senate and Democrats in the Assembly (Ward, 26). Although the lines must be approved by both houses, each majority gives the other a wink and a nod to assure each other’s survival. Unable to agree on a redistricting plan that could pass both houses and satisfy the courts, Assembly Speaker Stanley Fink finally proclaimed to Senate majority leader Warren Anderson, “You don’t quarrel with the way that I draw the Assembly and I won’t quarrel with the way that you draw the Senate. I will pass a bill that has your version of the Senate if you will pass the same bill that has my version of the Assembly” (Schneier, Murtaugh and Pole, 89). Aided by newly developed computer programs, both Assembly Democrats and Senate Republicans have become remarkably adept at drawing district lines to maximize their respective party advantages.

Under this cynical system, the legislators choose their voters rather than the other way around. The constitution that was built to check government’s power is reversed. The 59th Senate District is a fine example of such political engineering. Former Senator Dale Volker was a longtime incumbent and his district is based on partisan gerrymandering that favors Republican candidates. Reformers and public interest groups have long called for an independent redistricting commission to make legislative elections more competitive and make legislators more accountable to voters. A legislature-favored bipartisan commission on the other hand is more likely to engage in partisan gerrymandering with certain districts safely assigned to one party and other districts safely assigned to the other party. To remove partisan motivation from the redrawing of legislative district lines, numerous proposals have been advanced for vesting the responsibility in a nonpartisan redistricting commission.

Case Study: Upstate Prisoner Representation

Backtracking on previous campaign promises to commit to a nonpartisan redistricting plan, the Republicans abandoned the democratic ideals for political gain. When former Governor David Paterson signed a law banning gerrymandering of prisoner populations, the Republicans were outraged. Under the groundbreaking legislation, prisoners will be counted towards their home districts before incarceration and not the districts where the prison or detention unit is located. Crucial to Republican-controlled districts, the Upstate prisons account for a considerable population in counties that are otherwise under the minimum threshold. The practice of counting inmates as residents of the district grossly inflates the overall population of an electoral district. The practice of counting inmates where they are incarcerated has long been criticized as disenfranchising the poor, urban and largely Democratic minority neighborhoods by undercounting their populations, and thus reducing their political clout when legislative districts are drawn. The net impact of the new legislation will restore prisoner populations to Downstate.

In a recent article titled “New York Republicans Challenge Ban on Gerrymandering”, nine Republican State Senators filed a lawsuit challenging the ban on prison-based gerrymandering. One of the protesting Senators, Retty Little, has 13 prisons in her district each accounting for more than 10,000 prisoners from various parts of the state (Loftin). Government watch groups characterized the Republican tactic as a political obstruction to one of the greatest civil rights accomplishments of the last decade in the State. However, seeing that the Republicans are clinging on to a fragile majority over the Democrats in the State Senate, it is safe to say that they are guaranteed to fight any redistricting reform that could cost them seats that are considered safe for their party. Senator Joseph A. Griffio, another Republican whose district relied on its prison population, said that the change in law underscored how the Legislature was dominated by Democrats from New York City. Griffio noted that “This is another attempt by New York City leadership to silence the upstate voices” (Loftin).

While the Republican defense for gerrymandering is deplorable, ultimately the Upstate voters benefit from a Senate Republican redistricting plans that would benefit Upstate representation. It is clearly not in the interest of

Upstate citizens to lose political representations due to declining population. It is also not in the interest of the Republican Party to lose dominance over the State Senate with new redistricting plans. For that reason, the Republican Party and the Upstate residents have formed a bond that is the key to ensure both of their political survival by keeping to a system that has largely maintained the status quo.

Future Prospects

Long time journalist of Albany politics, Jay Gallagher, published a book-length explication of criticisms linking governmental failure to weakness in New York’s economy. Much of the blame, he wrote, “can be laid at the feet of the men and women we send to the Capitol to serve us. All too often, the politicians act in their own self-interest and that of the thousands of lobbyists who spent \$144 million last year trying to influence them and less for the general good” (Ward, 34). What Gallagher argues is that Albany seems to be lacking the principle that should be its focus: a dedication to the prosperity and wellbeing of its citizens. It also doesn’t help that it has been known as an institution that did most of its important work in the dark of night- with “lawmakers so exhausted or confused or just fed up that they would vote for almost anything just to end the pain. Merely ending the legislative session in daylight was considered something worth cheering” (Gallagher, xi). Under such conditions, how can the public even be sure that their best interests are being considered?

The current situation in Albany has yet to give voters compelling reasons to participate in the political process. The combination of strong partisanship and a divided government complicates and distorts the process. The Republican minority in the Assembly and the Democratic minority in the Senate continually complain that the majority party singlehandedly controls the legislative process, thereby making it impossible for minority members to have major bills supported and approved. The State Executive-Legislative relationship is also subjected to partisanship. The parties in majority of either house have huge bargaining power with the Governor. Frank J. Munger and Ralph A. Straetz explained that the New York State Legislature “possesses a party discipline far superior to that found either in the United States Congress or in all but a few capitals... Consequently, the process of legislation consists of a process of negotiations among the legislative leaders... and the Governor” (Gallagher, 105).

Revitalizing the Upstate economy requires a coordinated and sustained effort from all parties and all levels of the government. If the politics is a cause of the decline, then it also must be its savior. The politicians in Albany should have no higher priority than to stand up for New York State's economic future and making it the best place to do business in America. First, the business climate must be improved to stimulate economic development and create new jobs. Second, we must revitalize cities and downtowns to make them economically vibrant places to live and work, especially for the young population. With dedicated leadership, New York will emerge as a state where ideas are born, companies are raised and jobs and careers thrive. If those promises come true, and if a new Upstate generation can rediscover the entrepreneurial genius of the past, the region's prospects look bright.

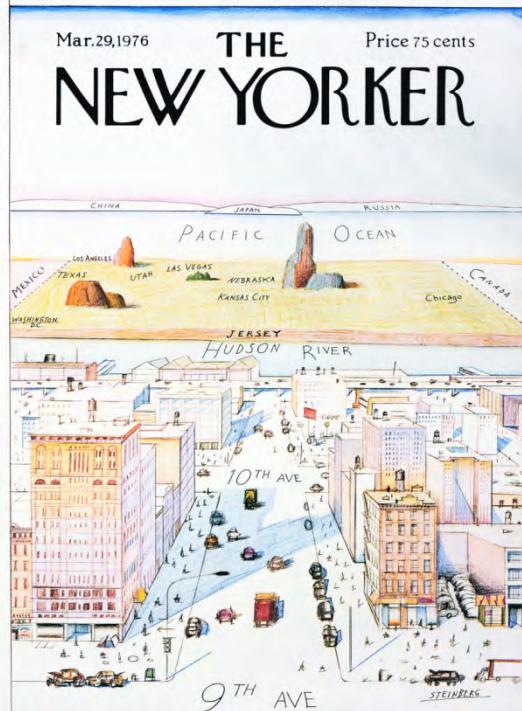
As for the 59th State Senate district, it appears that its prospects will remain the same in the near future. Without fair nonpartisan redistricting, Republican candidates will continue to dominate races as they have for decades. Without enticing job offers in the region, the young college graduates will continue to look for work elsewhere in the state and the nation. Without a vibrant pool of young and skilled workers, prospective employees and entrepreneurs will also look elsewhere to start their businesses. The cycle of economic stagnation will continue to perpetuate itself unless politicians from both parties are willing to cooperate on the only issue that should be on the top of their agenda- bringing meaningful legislation that will benefit all of New Yorkers.

While the Upstate-Downstate gap will continue to dominate political tensions, politicians

and constituents alike should all look past their differences and towards a common goal. Excessive partisan politics should especially be abandoned for the purpose of finding a middle ground that can be accepted for all citizens. If New York wants to truly be great again, the two regions must bridge their divide in Albany, and help each other with the necessary legislation even if it means agreeing to concessions they don't necessarily want to make. Albany should pay particular attention to Upstate concerns and not just cater to Downstate's powerful influence. While Republicans have found a political niche by using their Upstate stronghold to constantly foil the Downstate Democrats, perpetuating this stalemate will only prolong existing problems and inefficiency in the legislative process.

It will be interesting to see what Patrick Gallivan will do to improve the livelihood of Western New York. This senior thesis experience has really brought down my wall of resistance to pro-Republican notions. While by no means will I become a Republican anytime soon, I have decidedly erased my preconceived notions about the ills of the GOP. With the system we have now, Republicans are in fact helping Upstate citizens retain their share of representation in the state legislature against powerful Downstate interests. But I still hold on to the belief that if Cynthia Appleton were elected, an Upstate Democrat is just as capable as a Republican in representing Upstate interests. What politicians need to do in New York is to put aside party agendas and truly put their constituents' interests in mind. Representative democracy is not dead in New York as long as the citizens are informed and care enough to do something about their rights to be heard.

Figure 1



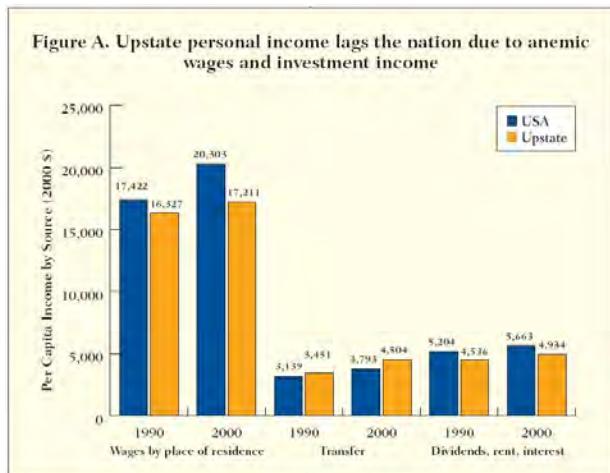
Source: *The New Yorker Magazine*

Table 1

59th SENATE DISTRICT												
County	Cynthia M. Appleton	Patrick M. Gallivan	Patrick M. Gallivan	Patrick M. Gallivan	Cynthia M. Appleton	David J. DiPietro	Blank	Void	Scattering	BVS	Total	
	DEM	REP	IND	COR	WOR	TEA						
Wyoming	3,686	5,248	289	688	610	1,836	523	0	0	523	12,861	
Part of Erie	14,538	30,447	2,745	4,545	2,636	7,549	3,412	0	0	3,412	65,870	
Part of Livingston	4,296	6,007	438	1,054	379	630	834	9	4	847	14,251	
Part of Ontario	4,259	5,764	385	1,019	338	835	1,067	1	3	1,071	13,641	
Total	26,739	48,067	3,037	7,324	3,653	10,830	5,938	10	7	5,933	106,623	
RECAP	30,712	59,208				10,830						

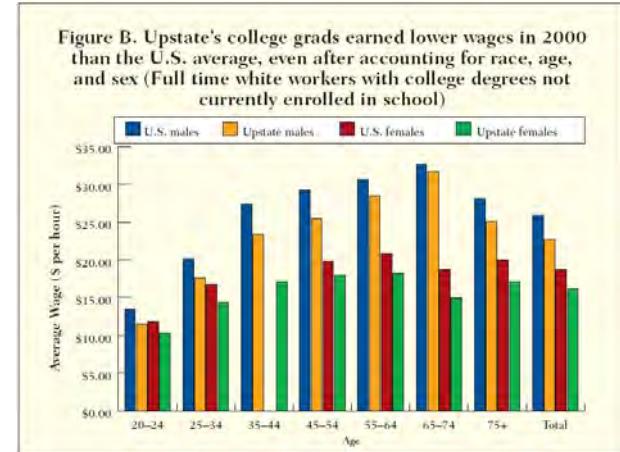
Source: *New York State Board of Election*

Figure 2



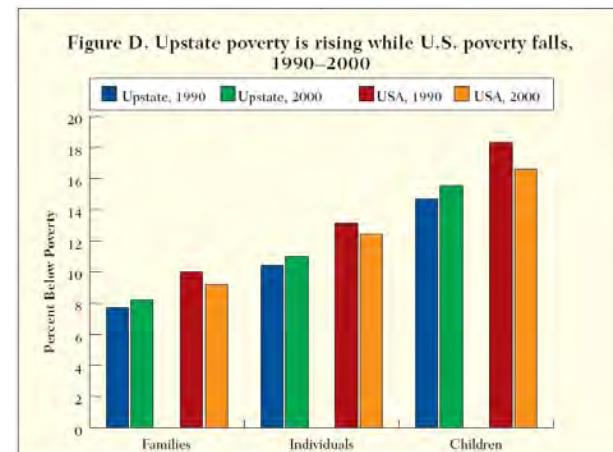
Source: *The Brookings Institution*

Figure 3



Source: *The Brookings Institution*

Figure 4



Source: *The Brookings Institution*

Works Cited

- Barrett, Devlin and Lisa Fleisher. "New York State Loses 2 Seats in House ." *The Wall Street Journal* 22 December 2010.
- Gallagher, Jay. *The Politics of Decline*. Albany, New York: Whitston Publishing Company Inc., 2005.
- Gardner, Kent. "Balance of Revenue & Expenditure Among New York State Regions." May 2004. *Center for Governmental Research*. 20 April 2011
http://www.cgr.org/reports/04_R-1400_NYBalanceOfPayments.pdf.
- Hunter, Judith. *Reflections on the 59th Senate District Race*, Derek Weng. Personal Interview 27 January 2011.
- Lind, John. "What on Earth Happened to the New York GOP? ." *The Wall Street Journal* 18 September 2010.
- Loftin, Britton. "New York Republicans Challenge Ban on Gerrymandering." *Politic 365* 14 April 2011.
- Metropolitan Transportation Authority. *New York City Transit: Facts and Figures*. 10 April 2011
<http://www.mta.info/nyct/facts/ffintro.htm>.
- New York State Board of Election . *Enrollment By County*. 29 March 2011
<http://www.elections.state.ny.us/EnrollmentCounty.html>.
- New York State Board of Election. *2010 Election Results* . 28 March 2011
<http://www.elections.state.ny.us/2010ElectionResults.html>.
- . *Campaign Finance Disclosure Reports*. 10 April 2011
<http://www.elections.state.ny.us/CFViewReports.html>.
- Office of the New York State Comptroller. "Analysis of Fiscal Stress in New York State's Cities." February 2006. 20 April 2011
<http://www.osc.state.ny.us/localgov/pubs/research/citiesfiscalstress.pdf>.
- Office of the State Comptroller. *The Citizen's Guide: A Pocket Reference to New York State's Finances*. Albany, New York: The Office of the State Comptroller, 2010.
- Pendall, Rolf and Susan Christopherson. *Losing Ground: Income and Poverty in Upstate New York, 1980-2000*. Washington D.C.: The Brookings Institution, 2004.
- Pendall, Rolf. *Upstate New York's Population Plateau: The Third-Slowest Growing "State"*. Washington D.C.: The Brookings Institution, 2003.
- Roberts, Sam. "The Upstate Economy, the Problem That Still Plagues Politicians and Their Promises." *The New York Times* 11 June 2006.
- Schneier, Edward V., John Brian Murtaugh and Antoinette Pole. *New York Politics: A Tale of Two States*. Armonk, New York: M.E. Sharpe, Inc., 2010.
- Seward, James L. *MTA Bailout Lets Down Upstate*. 15 May 2009. 29 March 2011
<http://www.nysenate.gov/blogs/2009/may/15/mta-bailout-lets-down-upstate>.
- Surtel, Matt. "Domagalski Withdraws from 59th State Senate District race." *The Daily News* 4 October 2010.
- The Drum Major Institute . *Solving the MTA's Budget Crisis and Reinvesting in Mass Transit*: New York: The Drum Major Institute for Public Policy, 2010.
- U.S. Census Bureau. *State and County QuickFacts: Livingston County, New York*. 4 November 2010. 26 March 2011
<http://quickfacts.census.gov/qfd/states/36/36051.html>.
- Ward, Robert B. *New York State Government* . Albany, New York: Rockefeller Institute Press, 2006.

Psychedelic Research Resurgence

Andrew Demosthenous

Introduction

In the 1960s, more than 1,000 peer-reviewed research articles detailed the therapeutic use of psychedelic compounds. Passage of the Controlled Substance Act of 1970 classified psychedelic substances such as LSD, MDMA and Psilocybin as Schedule 1 drugs with no medical value, interrupting research into the application of psychedelic compounds for well-being. New studies from institutions such as New York University, Purdue University, Harvard University, and Johns Hopkins University are producing evidence that use of these drugs in combination with psychotherapy may have potential value for treating mental illnesses such as addiction, mood, affective, and psychotic disorders. This study explores the relationship between the effects of these drugs on well-being and the FDA scheduling of these drugs at the highest level. Analyses of six open-ended interviews with a snowball sample of college students about their recreational experiences with LSD, MDMA and Psilocybin found that college students use these drugs as a tool for spiritual and emotional development. They reported positive transformative experiences and improved relationships with family and friends. These findings support the continued expansion of research on psychedelic assisted psychotherapy for treatment of mental illness and the rescheduling of medically beneficial psychedelics by the FDA.

Hallucinogens

“Hallucinogens are substances that cause powerful changes in sensory perception, from strengthening a person’s normal perceptions to inducing illusions and hallucinations” (Comer 2010, p. 389). These substances, also known as psychedelic drugs, include LSD (Lysergic Acid Diethylamide), Psilocybin (4-phosphoryloxy-N,N-dimethyltryptamine), and MDMA (3,4-methylenedioxymethamphetamine). Many people use different definitions for hallucinogens and include different types of drugs into those definitions. For the purpose of this paper, the hallucinogens being discussed will be the substances which affect serotonin (5-HT) type

receptors. Many hallucinogens are found naturally in plants and animals. Evolution of tools in chemistry and the scientific method in the 1900s supported the process of synthesizing new hallucinogens.

With the enactment of the Controlled Substances Act of 1970, hallucinogens, among them LSD and Psilocybin, were classified as schedule I drugs by the United States congress (U.S. Drug Enforcement Agency 2010). In 1984, MDMA was added by the Drug Enforcement Agency (Baum 1997, p. 213). The schedule of the drugs, I being the highest and V being the lowest, determines the regulatory measures enforced by the United States DEA. The higher the drug scheduling, the stricter the policies are regarding manufacturing, importation, possession, distribution, and research for that particular drug. Placing LSD, Psilocybin and MDMA in the schedule I classification disrupted research on the application of hallucinogens to well-being and mental health. This also fostered the strictest judicial consequences for possession and distribution. Both using and distributing LSD are felonies. Anyone caught distributing LSD can get up to forty years in a federal penitentiary and a two million dollar fine (National Geographic 2009).

Three criteria the United States DEA and FDA use for determining if a substance should be classified as schedule I are: 1) the drug has no accepted medical use in the United States; 2) there is a lack of safety for use of the drug under medical supervision; 3) the substance has a high “potential for abuse” (Nichols 2003, p. 133). “Potential for abuse,” an ambiguous criterion, is typically interpreted by the United States DEA and FDA as the potential for addiction.

The FDA system of scheduling has been heavily criticized due to the influence of politics. President Nixon in the early 1970’s took responsibility for drug scheduling away from the Surgeon General and his team of medical experts. The responsibility was delegated instead to the office of the attorney general and his team of law enforcement experts. Since then, “cops and lawyers instead of doctors were judging the toxicity of drugs” (Baum 1997, p. 25-26). Decisions made by the U.S. attorney general’s office regarding the

scheduling of drugs were more heavily influenced by the politics of the war on drugs than scientific data presented by doctors. For example, in the hearings held by the DEA in 1986 regarding the scheduling of MDMA, thirty-four witnesses, including experts in the field of psychology and psychiatry, gave testimony and a ten volume report outlining the potential medical benefits of MDMA, asking for its rank to be lowered to schedule III. DEA Administrative Law Judge Francis Young wrote a ninety page opinion paper highlighting the witness' findings and agreed that MDMA should be dropped to schedule III, IV or V because of its potential medical value in mental health. Despite the overwhelming evidence and the agreement by Judge Young that the law should be changed, Judge Young reinstated his original ban on MDMA, leading to the permanent placement of MDMA in Schedule I of the Controlled Substances Act in 1986 (Stafford 1992, p. III-63).

There is also evidence that the government was aware of this inaccurate evaluation of LSD's danger because the CIA had conducted multiple experiments on American and British military personnel. From 1947 to 1977 the CIA had conducted operations coded MKSEARCH, MKOFDTEN, MKCHICKWIT, MKULTRA, MKDELTA, and MK NAOMI (General Council of the Department of Defense 1977). Their goal was to use LSD as a covert weapon to cause brain damage or permanent insanity to the enemy. Their other goal was to use LSD as a mind control agent or torture device for prisoners in order to obtain information (Greenberg 2010, p. 164) (National Geographic Documentary 2009). It would have been the perfect weapon because a 25 microgram drop (the size of two salt grains) of this colorless and odorless fluid anywhere on the skin could cause the effects. One ounce is enough to dose 300 thousand people. The CIA was unable to achieve their goals of weaponizing LSD. They were unable to control minds or cause permanent insanity by dosing personnel. Instead of being controlled or damaged, the men had fits of laughter and an enjoyment for the nature environments they were in (National Geographic Documentary 2009).

Historical Uses of Psychedelics

Psychedelics occur naturally in both animals and plants. They have played a significant role in the development of philosophy and religious thought in earlier cultures. Well documented examples of psychedelic use include the soma of

ancient India, peyote used by Native American tribes, rituals by Aztec shamans, ritual use by peoples throughout Mexican history, ceremonies of the ancient Greeks, and use of ayahuasca in modern Brazil (Nichols 2003, p. 133). Psychedelics were popular for philosophy development and religious use because of their ability to induce states of altered perception. These substances were not considered "drugs" in these cultures. To find references to use in literature, it is important to look at key words such as "fruit of the gods," "divine fruit," or "flesh of the gods."

The science of psychedelics became popular after new scientific methods for synthesizing psychoactive substances became available and new techniques for studying them. MDMA was first synthesized in 1912 by Merck chemist Anton Kollisch (Ecstasy Rising 2004).

MDMA hadn't been ingested at that point, so there was no indication of its potential effects. Soon after Albert Hoffman synthesized LSD at the Sandoz pharmaceutical laboratory in 1938, scientific interest began to peak in psychedelics. Hoffman had accidentally ingested a small amount and experienced the world's first LSD trip (National Institute on Drug Abuse 2001). By the 1960s more than a thousand peer-reviewed research articles detailed the use of psychedelic compounds for medical purposes. Promising therapeutic effects were reported in over 40,000 subjects who had each used LSD in multiple sessions (Vollenweider and Komter 2010, p. 1) (Nichols 2003, p 162) (Stafford 1992, p. 40-44). At the same time, recreational use was rampant, having been made popular by Leary and his associates (Stafford 1992, p. 51). The suicide of Diane Linkletter, daughter of the entertainer Art Linkletter, after taking LSD was one of the major sparks that allowed the passing of the Controlled Substance Act of 1970 which put all psychedelics on the schedule I list. The media had argued that Diane had no personal problems, and it was the drug that caused the suicide. While alive she was portrayed as a well-educated, Christian, white woman with no prior drug use or symptoms of depression (Baum 1997, p. 25) (Greenberg 2010, p. 256). Rampant use of ecstasy in the 1980s lead to the adding of MDMA to schedule I in 1985. The Controlled Substance Act and the war on drugs interrupted all research until the 1990s, where research began to make a slow comeback.

A recent resurgence of research on these substances questions the validity of their schedule I classification. Media outlets such as *World Report* and the *New York Times* have expressed excitement

about the psychedelic assisted psychotherapies for the treatment of anxiety, depression, post-traumatic stress disorder, obsessive compulsive disorder, addiction, and more. Tom Roberts, professor of educational psychology at Northern Illinois University, notes that “What we see now is the [US] FDA making decisions based on data rather than politics” (Morris 2008, p. 1491). Universities, among them Johns Hopkins, Purdue University, University of Arizona, Harvard, New York University, and U.C.L.A, are obtaining federal permission to study these substances, leading the way in the United States. Non-profit groups such as the Heffter Research Institute and the Multidisciplinary Association for Psychedelic Studies (MAPS) are also contributing (Tierney 2010).

Research Supporting Lowering LSD, Psilocybin, and MDMA from Schedule I

Since the 1960s, public service announcements have been making false claims about LSD and Psilocybin that had not been confirmed by ethical scientific research, frightening drug effects ranging from schizophrenia, chromosome damage, birth defects, and murder to suicide (National Geographic Documentary 2009). False claims about MDMA have included irreversible brain damage, Parkinson’s after one pill, and holes in the brain (Ecstasy Rising 2004). Current leaders in this field, who have performed ethical clinical trials with animals and humans, disagree with these claims. Their research also shows that these hallucinogens meet none of the criteria required for schedule I designation.

1) The drug has a high potential for abuse.

Hallucinogens do not engender drug dependence or addiction and are not considered reinforcing substances. They don’t produce drug-seeking behavior or physical withdrawal symptoms (Nichols 2003, p. 134) (Studerus et al. 2010 p. 2). The scientific community argues that dependence liability in a drug stems from the ability to affect dopaminergic (DA) transmission. “Nearly all hallucinogens lack the affinity either for DA receptors or for the DA uptake transporter and therefore do not directly affect DA neurotransmission (Nichols 2003, p 134).” There are no literature reports of successful attempts to train animals to self-administer hallucinogens, indicating that these drugs do not possess the

necessary pharmacology to initiate or maintain dependence (Nichols 2003 p. 134) (Studerus et al. 2010 p. 14). Results of a clinical study involving 227 psilocybin sessions with 110 subjects conducted between 1999 and 2008, indicate that “carefully monitored administration of psilocybin to healthy volunteers within an experimental setting does not increase the risk of subsequent abuse of psilocybin or other illicit drugs (Studerus et al. 2010, p. 14).” These results are in line with another long term study involving LSD. In a ten year follow up of 247 subjects who received LSD in an experimental or therapeutic setting, most subjects reported discontinuing or reducing their hallucinogen drug use. Use of hallucinogens tends to be less frequent or discontinued over time. Regular use is unlikely for several reasons.

Tolerance develops at a rapid rate (Studerus et al 2010, p. 14). Hallucinogens, when taken for a second consecutive day, do not give the user the normal effects. Subjects in the psilocybin study reported the effects as being tired. They were glad to regain their normal state of consciousness and needed time to recharge. Furthermore, hallucinogens do not produce the pleasurable effects of addictive drugs such as escape, euphoria, anxiety relief, increase of self-esteem, etc (Studerus et al. 2010, p. 14).

2) There is a lack of safety for use of the drug under medical supervision.

“Hallucinogens are generally considered to be physiologically safe molecules whose principle effects are on consciousness (Nichols 2003, p. 134).” They do not have harmful physiological effects. Hallucinogens are not toxic to any mammalian organ systems and there is no evidence to suggest that they damage any human organs. “Hallucinogens do not cause life-threatening changes in cardiovascular, renal, or hepatic function because they have little or no affinity for the biological receptors and targets that mediate vital vegetative functions (Nichols 2003, p. 134).” Dr. James Gill, Deputy Chief Medical Examiner in New York City, conducted a three year study in the early 2000s to see how many deaths in New York City were related to MDMA toxicity. He looked at all the autopsies that were not related to any natural causes of death. Out of 19,000 autopsies, only 22 subjects had MDMA in their system. Of these 22 subjects, only two of the deaths could be attributed to ecstasy alone. This study was conducted at a time when statistics from

the DEA reported New Yorkers using approximately 110 million doses of ecstasy (Ecstasy Rising 2004).

The primary effect of administering hallucinogens to subjects is psychological. LSD and psilocybin can induce disturbing experiences. They can also catalyze an onset of psychosis or depression, which has led (in some cases) to suicide. However, these drugs do not appear to produce illness in people who are not predisposed to those illnesses (Nichols 2003, p. 135). Psychosis and depression were catalyzed only in people having a genetic predisposition. The actual numbers of occurrences for recreational users are also low. A search of Medline in 2003 by Dr. Nichols for cases of LSD-induced psychosis yielded three reports in the previous 20 years (Nichols 2003, p. 135). According to the 2006 National Study on Drug Use and Health, approximately 23 million people (9.5 percent of the population) over the age of twelve had used LSD in their lifetime.

Risks can be avoided in ethical and responsible medical settings. Bad trips can be treated with “talk-down” therapy and administration of benzodiazepines (Nichols 2003, p. 135). Data from the psilocybin study demonstrate safety and tolerability in both the short term and long term with 227 psilocybin administrations. There is no indication of persisting perception disorders, prolonged psychosis, or other long-term impairments of functioning. A small proportion of “bad trips,” were resolved by providing interpersonal support. The team concluded “that psilocybin administration to healthy, high functioning, and well-prepared subjects in a responsible clinical or research setting is generally well tolerated and that future studies using this important research tool are justified (Studerus et al. 2010, p. 16).” The administrations were well tolerated, producing positive outcomes. The team concluded that hallucinogenic drugs used under controlled and supportive conditions can lead to,

“Sustained positive changes in personality, attitudes and values, particularly in those subjects who have experienced profound personal insights and transcendent or mystical experiences. Among the most often reported are more self-understanding, more tolerance of others, less egocentricity, a less materialistic and aggressive orientation, and more appreciation of music, art, and literature (Studerus et al. 2010, p. 14).”

Much contradictory research that emerged during the war on drugs neglected two very important concepts regarding hallucinogens. The first is that hallucinogens are not predictable like other central nervous system drugs. They do not have the same effects every time because primary determinants of subject experience are expectations (“set”) and environment (“setting”) (Nichols 2003, p. 137) (Greenberg 2010, p. 164). Hallucinogens do not directly cause any behaviors or feelings. They amplify the thoughts and emotions of the user and are completely vulnerable the user’s “set” and “setting.” A negative “set” and “setting” can result in a “bad trip” or a disturbing experience. If a clinical or research setting creates a positive “set” and “setting,” the experience of the user can be sublime. Early research supporting the current scheduling of hallucinogens used a very negative “set” and “setting” for the subjects. Subjects were administered the drugs in very rigid, mechanical, and frightening settings. They were strapped down to a bed to be poked and prodded with tools (National Geographic Documentary 2009). The CIA also conducted unethical experiments by dosing military soldiers without their knowledge or consent. This led to a very small percent of soldiers committing suicide via jumping out of windows because they falsely believed they became permanently insane (Greenberg 2010, p. 164). It was the situation which elicited negative reactions from the patients, not the drug itself. These situations can be controlled. Secondly, different doses produce different qualitative effects (Nichols 2003, p. 137).

Increasing or decreasing dosage does not simply change the amplification of the effects. There are different thresholds for different tiered effects. For example, there is a certain dose required for people to hallucinate on LSD and psilocybin. If the dose is lower than that threshold, the subject will not have hallucinations. This is valuable because the patient can be given a dose that can provoke the emotional and perceptive effects without causing uncomfortable hallucinations. This concept was neglected in research aimed to see if hallucinogens could be safely administrated.

3) *The drug has no currently accepted medical use in the United States.*

LSD, Psilocybin, and MDMA each have their own unique medical applications. They have a cross tolerance and several shared effects, but their

subtle differences in chemical structure cause them to have unique effects, and thus a diversity of medical applications. LSD and psilocybin are known as Classical Hallucinogens. Classical Hallucinogens have a high affinity for serotonin (5-HT2A) receptors. MDMA does not have the same affinity and is not considered a Classical Hallucinogen. There are several potential medical applications for administration of these three hallucinogens to patients in order treat illness. The illnesses which might be effectively treated with hallucinogens currently have few highly effective treatments available which also include harmful side effects, thus necessitating this new innovative research. There is also valuable information that can be obtained about cognitive neuroscience by studying the chemical interactions in the brains of healthy subjects under the influence of Classical Hallucinogens.

Clinical Applications of Psychedelics

Psychedelic Assisted Psychotherapy

The concept for the technique of psychedelic assisted psychotherapies is completely new. Psychedelic assisted psychotherapy involves using hallucinogens to put patients in an altered mental and emotional state which would make a therapy session more effective. Dr Roland Griffiths of the School of Medicine at Johns Hopkins University describes it as,

“Not about taking psilocybin or other compounds multiple times. It’s about orchestrating, if you can, a single profound transformative experience; that then results in an unfolding of behavioral change over time (National Geographic Documentary).”

These sessions are only meant to happen once or very few times and yield results of positive long term behavioral change. The whole process has several phases. First, the subject must be fully diagnosed by a licensed professional. A review of the subject's background and diagnoses will need to be analyzed to see if the subject has the necessary prerequisites to safely take the drugs (Multidisciplinary Association for Psychedelic Studies 2005, p. 7). The subjects must then have intense preparation before the psychedelic assisted session (Vollenweider and Kometer 2010, p. 4). The therapist must form a strong therapeutic alliance, create a safe psychological and physiological space, and prepare the participant mentally for the experience. These can be

accomplished in approximately two ninety minute sessions (Multidisciplinary Association for Psychedelic Studies 2005, p. 9). The drug assisted session will be longer than a typical session. It will be multiple hours; depending on which drugs are used and how long the effects from the dosage is. After the session, it is beneficial for there to be drug-free follow up sessions. All three drugs of interest, along with other hallucinogens, have special qualities for assisting therapy.

MDMA Assisted Psychotherapy

Preliminary trials with MDMA assisted psychotherapy have had very high rates of success with treating posttraumatic stress disorder. “Posttraumatic stress disorder (PTSD) is a debilitating anxiety disorder characterized by re-experiencing, hyperarousal and avoidance symptoms, and is a major worldwide public health problem (Mithoefer et al. 2010, p. 1).” “Patients with PTSD are prone to extremes of emotional numbing or overwhelming anxiety, and often have a narrow window between thresholds of under and over-arousal (Mithoefer et al. 2010, p. 1).” According to the 2005 Multidisciplinary Association for Psychedelic Studies (MAPS) handbook,

“MDMA can attenuate the fear response and decrease defensiveness without blocking access to memories or preventing a deep and genuine experience of emotion. Participants are able to experience and express fear, anger, and grief with less likelihood of feeling overwhelmed by these emotions. MDMA seems to engender awareness that such feelings arise as an important part of the therapeutic process. In addition, feelings of empathy, love, and deep appreciation often emerge, along with a clearer perspective of the trauma as a past event and a heightened awareness of the support and safety that exists in the present. Hence, the goal of MDMA-assisted therapy in treating PTSD is to enable the participant to restructure his/her intrapsychic realities and develop a wider behavioral and emotional repertoire with which to respond to anxiogenic stimuli.”

Many therapists who have experience performing MDMA assisted psychotherapy claim that it could accomplish what would take years of therapy in one session (Baum 1997, p. 212) (Greenberg 2010, p. 154). An example of feedback from a patient is that the drug helped her

confront memories of a rape she suffered, which helped her end periods of depression in panic. She told *Newsweek*, “not only did MDMA enable me to recover my sanity, it enabled me to recover my soul” (Baum 1997, p. 212). MDMA assisted psychotherapy is primarily being used for PTSD, however it has also shown promise in couple’s therapy and mood and anxiety disorders. MDMA, while classified as a hallucinogen and psychedelic, does not cause psychotic-like states or hallucinations like LSD or psilocybin.

Possible theories about the chemical mechanism for MDMA are that MDMA releases serotonin, stimulates the serotonin (5-HT2) receptor, and increases the neurohormones oxytocin, prolactin and cortisol. The mood and perception lifts are caused by the serotonin release (Mithoefer et al. 2010, pg. 3). “Findings suggest that oxytocin is involved in affiliation, trust, and accurate perception of emotion, so elevated oxytocin might help participants form a therapeutic alliance and revisit traumatic experiences in an emotionally engaged state (Mithoefer et al. 2010, p. 3).” It was also found that oxytocin reduced activation of the amygdale in humans, which is responsible for fear emotions. A recent study also found that elevation of oxytocin was associated with greater sociability and gregariousness (Mithoefer et al. 2010, p. 3).

LSD and Psilocybin Assisted Psychotherapies

LSD and psilocybin have different effects than MDMA because they agonize different serotonin receptors. They both primarily agonize the serotonin (5-HT2A) receptor which plays a major role in their medical applications. Researchers believe LSD and psilocybin may be useful in treating depression, anxiety, and Obsessive Compulsive Disorder (OCD) as an adjunct to therapy. Several universities, including New York University, Johns Hopkins University, and University of California have had a lot of success with treating anxiety and depression in terminally ill cancer patients (Vollenweider and Komter 2010, p. 3). Studies from the 1960s concluded that LSD treatment resulted in improved psychological adjustment in dying patients, made them more responsive to their families and environments, and enhanced their ability to enjoy every-day life. Later studies at Spring Grove State Hospital in Maryland showed improvement in about two-thirds of the cancer patients who had received LSD. Some of these reported

improvements were improved mood, reduced fear of death, and reduction in the amount of pain-relieving medication required. Half of patients reported dramatic improvement and profound experiences (Nichols 2003, p. 163). Dr. Charles Grob and his assistant Alicia Danforth conducted an end of life study at Harbor UCLA Medical Center which involved giving psilocybin assisted psychotherapy to twelve terminally ill patients. Terminally ill patients can be overwhelmed with anxiety and fear about their upcoming death. They also are typically on a lot of pain medication and are in medical environments which restrict them from appreciating life. Before dying, the patients reported positive mood shifts and increased acceptance of death. They had mended relationships and were able to spend quality time with their families (National Geographic Documentary 2009).

There are several theories for possible mechanisms to explain how classic hallucinogens improve therapy. Scientists estimate that the brain takes in eleven million bits of information per second; however, the brain can only process about two hundred bits at a time. To handle that amount of information, the brain must filter out information that is not critical. It does this by connecting new information to preconceived concepts so that it does not have to analyze all of it (National Geographic 2009). Dr. Torsten Passie’s “Hollow Mask Study” at the Hannover Medical School in Germany concluded that schizophrenics and subjects intoxicated by LSD do not process information in that same way. The filters are lifted and the ability to connect new information to preconceived concepts is temporarily suspended. This means psychedelics temporarily de-condition the mind, resulting in the ability to take in more information, more sensory impression, more emotional impression, more visual impression, and more access to parts of the mind (National Geographic Documentary 2009). This allows the LSD user to approach problems in a more genuine and open minded way that neglects socially conditioned behaviors or roles. The increased mind power and stimulation comes from the increased activity of the serotonin (5-HT2A) receptor.

Activation of serotonin (5-HT2A) receptors initiates complex chemical reactions in the pre-frontal cortex. LSD and psilocybin have such a similar shape to serotonin molecules that they are able to fit in the receptor like a lock and key. Serotonin (5-HT2A) receptors are located on the

large glutamatergic pyramidal cells in the deep cortical layers. The chemical reactions from binding with the serotonin (5-HT2A) receptors releases glutamate across the synapse, which then activates the AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) and NMDA (*N*-methyl-d-aspartate) receptors on the cortical pyramidal neurons. In addition, LSD and psilocybin directly activate the serotonin (5-HT2A) receptors located on cortical pyramidal neurons. The synergy of activating the serotonin (5-HT2A), AMPA, and NMDA receptors on the cortical pyramidal neuron results in the increased expression of brain-derived neurotrophic factor (BDNF) (Gonzalez-Maeso and Sealfon 2009, p. 227) (Vollenweider and Kometer 2010, p. 5). BDNF is a chemical that promotes the growth and survival of neurons. A deficiency in BDNF may impair the health of neurons and lead to depression (Comer 2010, p. 250). Since serotonin (5-HT2A) receptors are linked to the prefrontal cortex, this means Classical Hallucinogens lead to increased neuron strength in the prefrontal cortex from the BDNF. Areas in the prefrontal cortex are responsible for complex mental functions such as cognition, mood, perception, self control, and somatic awareness (Nichols 2003, p. 162). The increased effectiveness of therapy may be derived from a boost of brain metabolism and strength in these mental functions.

Administration to Patients: LSD and Psilocybin for Treatment of Cluster Headaches

John Halpern's research, at the McLean Hospital at Harvard Medical School, illustrates that LSD and psilocybin may have a high success rate for reliably treating cluster headaches. Cluster headaches are severely painful and debilitating headaches that reoccur. Scientists do not fully understand what causes cluster headaches, but they know it is related to blood vessels, nerves and neurochemistry. There is no cure and few treatments. The available treatments, including oxygen and/or medication, are minimally effective. The mechanism is unknown, but LSD and psilocybin seem to be highly effective in relieving or eliminating cluster headaches. A survey by Halpern indicated that for 53 subjects, psychedelics prevented or stopped cluster headaches more reliably than medication. Halpern and his colleagues at the Hannover Medical School in Germany launched a clinical trial on six subjects with cluster headaches. The team developed a non-psychedelic form of LSD, LSD-BROMO, and

administered it three times over a fifteen day period. For five of the six participants the headaches were gone. All of the participants reported that their quality of life had totally changed (National Geographic Documentary 2009).

The addition of a BROMO molecule to LSD creates a molecule that is too big to fit in the receptors that would normally trigger intoxication. LSD-BROMO successfully relieves cluster headaches in the same way other LSD molecules do, except it has no hallucinogenic properties. Despite displaying no hallucinogenic properties, LSD-BROMO was added to schedule I and is illegal in the United States; thus restricting research and clinical trials. Since This trial, no further clinical trials have been funded to administer LSD, psilocybin, or non-hallucinogenic LSD-BROMO to treat cluster headaches (National Geographic Documentary 2009).

Understanding Cognitive Neuroscience through LSD-Like Intoxication in Healthy Subjects

LSD-like hallucinogens, including psilocybin but not MDMA, are psychotomimetic at high doses; meaning they induces psychotic like episodes (Gonzalez-Maeso and Sealfon 2009, p. 2) (Nichols 2003, p. 162). There are a wide variety of hallucinogens that mimic different types of psychotic symptoms and disorders. The effects are contingent on how the molecules are shaped and by what receptors they either agonize or antagonize. LSD in particular mimics the positive symptoms of schizophrenia (hallucinations and thought disorder). Psychotomimetic hallucinogens can be used as tools for understanding neuronal basis of psychiatric disorders, understanding a psychotic experience, and finding a cure for those psychiatric disorders (Vollenweider and Kometer 2010, p.1). Franz Vollenweider uses neuroimaging technology at the University Hospital of Psychiatry in Zurich to map brains on psychedelics. His data shows that a brain on hallucinogens looks the same every time (National Geographic Documentary 2009). This means clinical trials with psychotomimetics are reliable to be analyzed and compared for conclusions.

Schizophrenia is very difficult to treat and learn about because some of the symptoms make fruitful communication between the patient and doctor impossible. Formal thought disorders, loose associations, derailment, poverty of speech, avolition, and social withdrawal make verbal communication very difficult (Comer 2010, p.

462). LSD-like hallucinogens can help researchers and doctors learn more about the schizophrenic experience by taking it themselves or performing clinical trials. Once the subject is sober, they will have their previous mental capacity and intelligence restored and be able to articulate their experience in a more accurate way than a patient with a permanent psychopathology. Dr. Charles Nichols, Geneticist from the LSU Health Sciences Center, also believes “by understanding how LSD produces its effect at the molecular and genetic level, we could potentially understand mechanisms underlying diseases like psychosis that have very similar overlapping behaviors (National Geographic Documentary 2009).” By understanding the neurochemical and genetic causes of psychosis, psychopharmacologists could potentially be able to create drugs that negate those causes in order to treat psychosis more effectively.

One possible problem with finding treatments for certain psychotic disorders, such as schizophrenia, is that they are uniquely human disorders (Gonzalez-Maeso and Sealfon 2009, p. 1). This means there are no animal models to do tests on. Animal test subjects are crucial because scientists can perform tests on them that would be unethical for human subjects. They can be bred to control for variables and provide a larger supply of test subjects. Hallucinogens fix this problem because they give opportunities to test schizophrenic-like mice. According to Dr. Nichols, LSD induced mice display human symptoms of schizophrenia. This gives him the ability to study brains of schizophrenic mice and test potential treatments on them (National Geographic 2009). They can induce mice with schizophrenic episodes and test anti-psychotics to see if they bring the mice back to a normal state of health.

Hallucinogens agonize and antagonize various serotonin receptors, among other neurotransmitters. The serotonin receptors in particular that hallucinogens interact with affect complex parts of the human brain, including much activity in the prefrontal cortex and interconnecting pathways between the cerebral cortex and limbic systems (brain regions implicated in pathophysiology of many mental disorders) (Gonzalez-Maeso and Sealfon 2009. p 228). Pathology relating to mood and anxiety is associated with an imbalance or abnormal functioning of serotonin chemical reactions. Thus, testing and understanding the mechanisms of hallucinogens may lead to important clues for the basis of psychosis. The results obtained from the mechanism of Classical

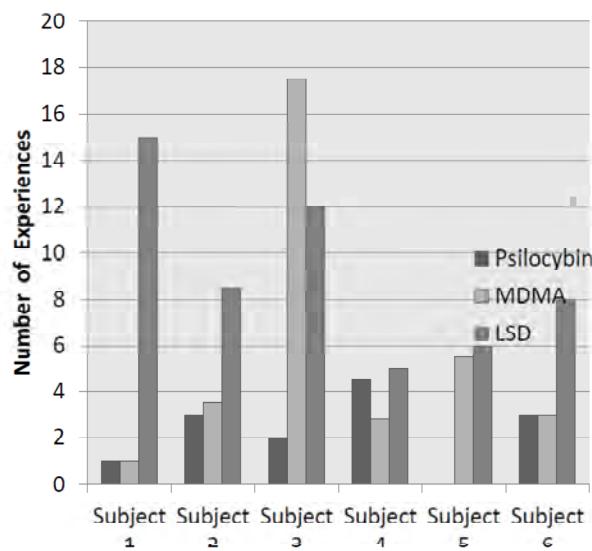
Hallucinogens may provide efficient approaches for the rational design and testing for new types of antipsychotic drugs (Gonzalez-Maeso and Sealfon 2009, p. 225). This also means experimental use with hallucinogens can be a great aid for mapping out the brain and further understanding the prefrontal cortex, pyramidal cells, serotonin, glutamate NMDA, AMPA, and BDNF interactions.

Methodology

Semi-structured interviews were conducted with a snowball sample of subjects to learn about their experiences with MDMA, LSD and psilocybin. The subjects were recruited privately through networking within the drug using community. Six public liberal arts College students were selected because of their high level of experience with the substances of interest. Five of the subjects were male and one was female. The goal of the interviews was to evaluate if the effects of MDMA, LSD, and psilocybin were recreational or entheogenic. The operational definition for “recreational experience” is experiences that cause altered physical and mental effects during intoxication. However, after the substance has been processed and excreted from the body, the subject retreats to their previous physical and mental states prior to intoxication. The operational definition of “entheogenic experience” is long term transformative effects on personality, spirituality, cognitive state, emotional state, and perception of one’s self after intoxication.

The interviews ranged from twenty minutes to an hour in length. The interviews were conducted at each subject’s home for comfort and privacy. To prevent coding bias, the interviews were recorded on a handheld recorder and later transcribed. The subjects gave verbal consent on the recorder. The recordings were deleted after transcription to protect the subject’s from being identified by their voice. No names or information that could be used to identify the subjects were used. The interviews had a pre-set list of questions; however, the subjects had freedom to discuss details beyond the scope of the pre-set questions. The subjects were asked questions about three different phases related to their drug use. They were asked questions about their experiences leading up to their first time uses with these drugs. They were asked questions about the experiences and feelings while intoxicated by these drugs. Finally they were asked about their experiences and reflections after their intoxication had ceased.

The subjects had varying levels of experience with each drug.



Bar Chart: Subject's estimate of experiences using Psilocybin (left bars), MDMA (middle bars) and LSD (right bars). Each subject had at least twelve combined experiences with the drugs of interest. All of the subjects reported doing some type of research before having their first experience with these hallucinogens. They approached these drugs with skepticism and caution because of the negative stigma associated with them. Each subject used sources such as books, encyclopedias, friends who had experience, and online sites for information about what to expect and how to approach the experience. All of the subjects mentioned using erowid.com for anecdotes, information about appropriate doses, and advice for activities to do while intoxicated.

Subject 5 (female) appeared to have recreational experiences while the rest of the subjects appeared to have, for the most part, entheogenic experiences. This can be explained by their different habits of use. Subject 5 typically took lower than average drug doses, which since hallucinogens have different qualitative affects at different dose levels, gave her less intense effects compared to the other users. She used the drugs to make regular recreational activities, such as dancing to music, more enjoyable.

Subjects 1,2,3,4, and 6 used hallucinogens differently than subject 5. They took higher doses and created their own private ritual-like experiences. They tended to avoid large social situations; instead, they spent their experiences in

small groups and in environments where they felt safe and comfortable.

Effects and Experiences While Intoxicated

Subject 5 used MDMA and LSD as party drugs. She used them to enhance social experiences. It enables her to connect with people more easily and with more depth. “On ecstasy I just love everyone.” The euphoria also amplifies the pleasure of music, colors and dancing. She also reported a greater connection with nature on LSD and psilocybin. The effects of the hallucinations were not as intense due to the relatively low doses she used. She described her environment as being “wavy” and “breathing.” Since the effects were not too intense, they were always very manageable. She claimed to feel very safe and confident in her drug use.

Subjects 1,2,3,4, and 6 had different experiences than subject 5. Their experiences were more intense and more varied. They reported having few negative moments in their experiences, but they were always able to calm themselves or each other down during those unfavorable intense moments. Aside from those few moments, they report overwhelmingly positive transformative experiences. They were very introspective.

According to subject 4,

“Whatever your feeling becomes dramatic and exaggerated and richer than you've ever experienced before. You kind of see your dark side and light side. You see your persona for what it is; for all its sincerity and bulls***. You see your shadow for what it is. You see your entire personality almost from an outside perspective and I think part of being on acid is losing that sense of judgment of positive and negative and ultimately coming to the conclusion that this is who you are. And as far as today goes, you're not going to change it. So you might as well embrace it.”

Subject 4 also reported the “[he] dance[s] across the entire emotional spectrum, but [he] does it with the volume turned all the way up.” He later said that he has experienced a range from ecstatic bliss to depressive nihilism.

Subject 1, when talking about LSD reported, “The way I experience the world is that everything is more beautiful. I'll just see something like a leaf with a dew drop on it and just love the natural beauty of it. Everything around just seems more vivid and I'm

generally pretty happy. I've never had a bad trip."

Subject 1 takes SSRIs, however, he reported a better mood lift with acid. "If I could take acid once a month, I could be content with the whole month."

The entheogenic subjects not only experienced introspective self-analyses, but also examined their relationships with family and friends. They reported that hallucinogens have been important aids in repairing or strengthening relationships. Four out of the five entheogenic users had specific anecdotes on how hallucinogens played a positive role in their relationships. Subject 1 reported,

"A couple of years ago I was having a fight with my girlfriend and I just took a couple of tabs of acid, probably not the best idea to do in that situation. I just went alone and did acid with a friend. It made me feel like, why am I fighting with my girlfriend? I should just go and make amends because it's stupid. I kind of felt love for everybody."

Subject 2 had an elaborate intimate experience with his significant other which he attributed to MDMA. They both took MDMA at his house.

"She was nervous at first. As we were coming up I comforted her and she became more comfortable with the sensation. From that point until we came down it was just heaven. We laid around in my room in our underwear and we just talked. We talked about each other and who are some people important in our lives. The whole experience centered around her and I and we really connected. And there are things that people say to each other, or don't say to each other rather. Like when you say very emotionally charged things to somebody you care about; sometimes you're apprehensive about overstating your appreciation for them to avoid maybe cheapening the relationship or feeling that you have for them. I think MDMA really let us say these things to each other and really mean it and understand that the other meant it. Like I said, sometimes you avoid saying these things to avoid cheapening what they mean and other times you simply don't get around to saying it them due to circumstances or timing or whatever. It let everything out that needed to come out. It was really instrumental in us two making that connection."

Subjects 3 and 6 talked about psilocybin experiences that elicited strong positive emotional feelings about family members. Subject 3 reported,

"One of my first experiences with psilocybin was at a concert and I remember enjoying the music very much. But at one point in the concert, I had just stopped and found my mind wondering and daydreaming. I was thinking about the relationship I had with my mother at the time. I didn't particularly like the way the relationship was going and so in the calm state I was in I wondered, Why isn't that better? Why don't I better that relationship?"

Subject 6 detailed one of his psilocybin hallucinations and emotions associated with it.

"My next journey was with my girlfriend. We were dressed like Adam and Eve in a forest. We held each other and spiraled into the green and blue corners of space and time. It wasn't erotic. It was a spiritual connection. I felt close to her. I felt one with her. I thought about my family. I connected with their spirits. Our energies became one. There was no more sadness. All problems ceased to exist. I met my father and we were free."

Post Experience Reflections

There were several common reflections among the subjects. All of the subjects feel safe, secure and confident with these drug experiences. They are all open-minded to using these substances again in the future. They are so confident in their experience and understanding of the drugs that they would feel comfortable recommending use to a friend or loved one. All of the users also reported having no regretted experiences on these substances. They felt the negative moments had value to their development and character.

The entheogenic users also had some common sentiments. They felt that their experiences with these drugs had a very positive impact on their relationships, mood, outlook on life, and creativity. They use these drugs primarily for spiritual and emotional development. They did not use frequently; rather, they decided to use at key moments in their life where they were struggling, had big decisions to make, or wanted to bond with friends.

Conclusion

Hallucinogens have a long history of being used by humans for spirituality and well-being. The act of moving hallucinogens to schedule I with the Controlled Substance Act was hasty and poorly thought out. This decision interrupted much needed

research into mental health and well-being. Current projects offer data that supports the lowering of hallucinogens from schedule I to a level that would continue to prohibit recreational use, while permitting research and medical use. There is overwhelming evidence that hallucinogens do not meet the criteria set by the Controlled Substance Act to be ranked as schedule I. Hallucinogens have a very low potential for abuse. They are one of the least addictive drugs that exist. They can be safely used in a medical situation because of a lack of toxicity and addiction. With the appropriate dose, set and setting, a licensed professional can create a safe and controlled environment for patients to use in a medical situation. The literature also shows that there are many potential medical uses in the United States from psychedelic assisted psychotherapy for the treatment of mood and anxiety disorder and Classical Hallucinogens for the treatment of cluster headaches, to clinical trials to aid in the creation of more efficient medications.

The research is in its preliminary stages and has not had many large trials. Much work needs to be done on understanding the mechanisms for how the drugs affect the brain and behavior at the molecular level. Success of preliminary studies justifies larger scale double blind trials. Results justify increased access to funding and permission for research institutions to use these substances in clinical trials.

The interviews support the literature and illustrate that it may be possible to orchestrate possible healthy and positive experiences with these drugs. Recreational users may not always give accurate insight or judgment about these drugs because street drugs are un-regulated. The drug's purity is unknown to the recreational user. Despite this flaw, the subjects interviewed did display characteristics predicted by the literature. Subject 5's description of how she feels on LSD is in line with Dr. Passie's conclusions from the "Hallow Mask Studies." It supported the mechanism by which LSD takes off the filters and allows a person to evaluate themselves or a situation in a new way. This account also illustrated the significance of "set" and "setting" because subject described how he feels a whole range of emotions instead of the same ones every time. All of the subjects displayed how hallucinogens can be a positive influence on mood. The anecdotes from Subjects 1, 2, 3, and 6 illustrated how hallucinogen helped them in their relationships. This supports the hypothesis that psychedelic assisted psychotherapy may be effective in couples therapy or family therapy.

Research into techniques for psychedelic assisted psychotherapy is imperative because they may be faster, safer and more effective than alternative methods used today; such as anti-depressants and other medications. According to Greenberg, anti-depressants fail to outperform placebos in clinical trials nearly half the time. They appear to make a positive difference in approximately 60 percent of the people who take them, while having serious side effects such as suicide (Greenberg 2010, p. 8). Mental health patients need newer treatments that are more effective, just as fast and without the serious potential side effects that exist with current medications.

References

- Baum, Dan. 1997. *Smoke And Mirrors: The War on Drugs and The Politics of Failure*. NY: Little, Brown & Company.
- Comer, Ronald J. 2010. *Abnormal Psychology: Seventh Edition*. NY: Worth Publishers
- Gonzalez-Maeso, Javier and Stuart C. Sealfon. 2009. "Psychedelics and Schizophrenia." *Trends in Neuroscience*. 32(4): 225-232
- Greenberg, Gary. 2010. *Manufacturing Depression: The Secret History of a Modern Disease*. NY: Simon & Schuster.
- Jennings, Peter. (Editor). 2004. *Peter Jennings Reporting: Ecstasy Rising* [Motion Picture].
- Mithoefer, Michael C, Mark T Wagner, Ann T Mithoefer, Lisa Jerome, and Rick Doblin. 2010. "The Safety and Efficacy of 3,4-Methylenedioxymethamphetamine-Assisted Psychotherapy in Subjects with Chronic, Treatment-Resistant Posttraumatic Stress Disorder: The First Randomized Control Pilot Study." *Journal of Psychopharmacology* 0(0): 1-14
- Morris, Kelly. 2008. "Research on Psychedelics Moves Into the Mainstream." *World Report* 371: 1491-1492
- National Institute on Drug Abuse. 2001. "Hallucinogens and Dissociative Drugs." *Research Report Series*.

- National Survey on Drug Use and Health. 2008. "Use of Specific Hallucinogens: 2006." *The NSDUH Report*.
- Nichols, David E. 2004. "Hallucinogens." *Pharmacology and Therapeutics* 101: 131-181
- Ruse, June M, Lisa Jerome, Michael C Mithoefer, Rick Doblin, and Elizabeth Gibson. 2005. "MDMA-Assisted Psychotherapy for the Treatment of Posttraumatic Stress Disorder: A Revised Teaching Manual Draft." *Multidisciplinary Association for Psychedelic Studies*. 1-51.
- Schedules of controlled substances. Retrieved from the U.S. Drug Enforcement Administration:
- Stafford, Peter. 1992. *Psychedelics Encyclopedia: Third Edition*. CA: Ronin Publishing Inc.
- Studerus, Erich, Michael Kometer, Felix Hasler, and Franz X. Vollenweider. 2010. "Acute, Subacute and Long-Term Subjective Effects of Psilocybin in Healthy Humans: A Pooled Analysis of Experimental Studies." *Journal of Psychopharmacology* 0(0): 1-19
- Tierney, John. 2010. "Hallucinogens Have Doctors Tuning In Again." *New York Times*. April 11. <http://www.nytimes.com/2010/04/12/science/12psychedelics.html?hp>
- Vollenweider, Franz X. and Michael Kometer. 2010. "The Neurobiology of Psychedelic Drugs: Implications for the Treatment of Mood Disorders." *Nature Reviews Neuroscience*. AOP: 1-10
- Wells, P.C. (Director). 2009. *National Geographic Explorer: Inside LSD* [Motion Picture].

Somatic Experiences and the Source of Religious Conviction

Joel Inbody

Why is it that religious believers accept the validity of doctrines that seem outlandish and strange to outsiders? Religion is truly a simple phenomenon, when it is broken down to its simplest form: a set of beliefs concerning what is sacred and profane and how one must behave accordingly. Despite this unifying Platonic idea, the various forms of religion are an extremely variable and widespread phenomenon among humans that cannot be as easily accounted for. The simplest tribes and peoples, such as the Aborigines of Australia, have fetish or totem objects; more complex societies tend to have institutionalized specialists who instruct the faithful in more complex dogmas and practices. Each physical representation is different and has characteristics so strange and alien that continuity between certain systems would seem mere contrivance. Yet each finds itself with followers who embrace the teachings and find truth in them. One must take into account the cultural and social influence, that is, the context in which the religion dwells. Tradition and training certainly play a great role in training up new generations of the faithful, as does the text or teachings found in each system. May there not be, however, a more natural element behind why such a multitude of humans, regardless of background and culture, place their trust in some religious faith? It seems strange that the human mind will merely accept the words of their elders without further reinforcement. Moreover, can any one source be touted as directly creating faith, or is belief produced by the interworking of multiple cogs and gears? While religion can be found on practically every inhabited landmass, an encompassing explanation for this phenomenon still remains elusive and transitory.

According to sociologist Clifford Geertz, the ritual itself is the source of religious conviction. Says Geertz, "...It is in ritual--that is, consecrated behavior--that this conviction that religious conceptions are veridical and that religious directives are sound is somehow generated... In a ritual, the world as lived and the world as imagined, fused under the agency of a single set of symbolic forms, turn out to be the same world." (Religion as a Cultural System) It is through the ritual that participants gain confidence that their beliefs are accurate depictions of reality. Religion, in Geertz'

words, is a set of cultural symbols that unite and personify a group of unique preferences. These preferences make up a cultural system that educates the performer concerning the values and norms that are acceptable and unacceptable in this social body. Without some means to prove these things, they would stand amongst and equal to others of differing sentiments and possibly opposite opinions. The purpose of acting out a ritual, then, is to make one particular set of preferences become true to the performer; it is to make the subjective, objective.

Though Geertz has clearly identified the ritual as the source of religious conviction, he is unclear as to how exactly it works to generate a belief in the performer's mind. Is it in seeing all his or her fellows carrying on in the same way that the neophyte understands, because he or she knows their companions and trust them, "This is true, this is what we should be doing"? Is it by acting in strange and unique manners that the reveler realizes the specialty of the event; that it is the culmination and self-fulfillment of their instruction? Or, rather, is it by some physiological means that this occurs, by some feeling called up and invoked for the occasion through an inherent something within the music; within the words on a page filtered through the mind; within the movement of the individual's body? Geertz does not elaborate, simply stating that conviction is "somehow generated" by these "acts of religious observance." Where, then, within the context of the ritual, does religious conviction emanate from? What is the actual mechanism? A ritual by itself can be very simple or very complex; it can be as basic as praying to a god with folded hands or as intricate as a dramatized performance on Cumorah Hill. It would be elucidating, then, to attempt to reveal just what it is that a ritual contains that inspires confidence in a set of sacred values. May it be that there are rituals that do not work because they lack certain components? Perhaps the entire process is subjective and varies based on the tendencies of the group and the individuals therein. Geertz' theory, while helpful in providing a specific area of focus, leaves considerable ground untouched in discovering the wellspring of certainty.

Within the context of the ritual, somatic experiences may provide the evidence required for

individual participants to accept the doctrines of that particular religion. These sensations, which seem to be provoked by music and the mental grasping of ideas, can give great internal satisfaction and meaning to the person experiencing them; they have a tactile character and bring stimulation to the mind and much of the body as well. Such sensations seldom occur and are unusually pleasant. Because of these factors they could be interpreted as, based on the context, a sign of spiritual activity or the specialty and sanctity of a performance. They are also intensely intimate and personal: somatic experiences seem to be brought on by some physiological reaction within the brain in connection with the conscious workings of the mind. For this reason, it seems safe to think that they may be experienced by all human beings regardless of their location or upbringing. If this is the case, cultural and social factors would work cohesively with the internal biology of an individual in creating religious conviction. A person who had rarely or never been exposed to somatic experiences in a secular setting may judge that responsible rituals are endowed with some spiritual energy or sacred fire. Somatic experiences, when encountered within the context of a ritual, may give the performer religious conviction.

Methods

Four individuals from Crossroads Christian Church in Elma, New York were interviewed concerning source(s) of religious conviction and the role somatic experiences might play in them. These persons have been dubbed Moses, Eve, Daniel, and Abraham, for the purposes of convenience and organization. The church they attend (usually referred to as simply "Crossroads") is self-described "non-denominational" though it is Protestant and Evangelical in doctrine (main beliefs include the necessity of personal salvation through the Sinner's Prayer and the Bible's inerrancy and sufficiency for Christianity.) It fits neatly into the recently devised category of neo-charismatic as described by (Robbins, 2004) and (Wuthnow, 1998). That is, it contains aspects of the charismatic and Pentecostal movements while eschewing institutionalization of these behaviors. For example, "worship" (a time of music and singing) is often energetic and individual, with a fairly even mix of participants who dance and wave their hands about while others simply stand and sing. Congregants can be infrequently heard praying in tongues under their breath, though this is a private act that will

receive neither praise nor admonition. In the service, great emphasis is placed on emotion, though it is restrained and kept in check by the music itself as well as the order of the program. At services, people talk openly about spiritual events that they have experienced. The general demographic of Crossroads is white, middle class, and middle-aged. There is also a substantial amount of the elderly, children, teens, and young adults. Minority groups, such as African-Americans, are either absent or are minute percentages of the population. The populace is mainly made up of those who live in the suburbs and outskirts of the city of Buffalo, which includes the towns of Elma, East Aurora, West Seneca, Holland, etc. About 100-200 people attend Crossroads each Sunday morning. A point was made to ask individuals for interviews with whom a pre-existing relationship existed and whom had spoken publicly about related issues.

The decision to focus on somatic experiences was based inductively on comments made during past church services. While preaching last winter, a Pastor at Crossroads once remarked that people have told him that when God speaks to them "they get goose-bumps." Additional comments have been made by congregants that seemed to imply that God's presence could be drawn from emotional experiences, or through music, and that he acted "inside." Tentative parallels could be drawn between the first example of "God speaking" by way of goose-bumps and something this researcher has dubbed the "realization sensation." The realization sensation involves an experience where someone who is contemplating an idea, whether in written or spoken form, makes a connection or realizes something. When this occurs, mentally, there is the onset of the awareness that something makes perfect sense; a concept or idea is incredibly clear and has been grasped like nothing else. Physically, a sensation of goose-bumps begins. It moves down the body from the top or middle of the head along the neck towards the shoulders. It is an objective sensation and can be clearly discerned when it occurs (the interpretation of its meaning, of course, is up to the individual.) This researcher has experienced it several times. At the earliest point, it was triggered while reading the Bible when a connection was made between the name of an angel in Revelation and a character in *Pilgrim's Progress*. It has also been elicited under a secular context, such as while studying Durkheim and collective effervescence. Such a potent perturbation may be

interpreted by some as God speaking or the presence of the Holy Spirit.

Questions were designed to gain information concerning somatic experiences and their possible role in religious conviction. These experiences included the realization sensation, but also encompassed tingles brought on by music as well as other peculiar happenings. The role of more traditional sources of conviction, such as the Bible, was also explored for each subject. Other questions were included with the purpose of discovering whether or not (assuming somatic experiences were considered evidence of spiritual things) this was due to socialization within the group. A sampling of questions appears below:

- Can you remember the first time you experienced or remember a sense of conviction? Was there some event or occurrence that caused you to really truly believe?
If you were raised a Christian, were you raised among the congregation of Crossroads, whether here or at a different location? (If no) What church or body?
- How important would you say that the Bible is to your faith?
- I often hear people at Crossroads talk of a personal relationship with Jesus, or walking with God; do you believe in such a thing? (If yes) What does this mean for you personally?
- Have you ever had an experience where you believed that God was directly communicating with you or leading you to do something? (Assuming the answer is yes) Could you please describe it to me?
- Do you believe that God can speak to you through the Bible? (Assuming the answer is yes) Have you ever felt that he was? How did you know?
- Do you believe that the Holy Spirit resides in you? (If yes) What evidence does he give you of his occupancy?
- Have you ever felt a tingling sensation up and down your spine, or along your skin, or in your head, during the praise and worship? (If yes) What do you think that means?
- During worship time at church, I sometimes observe people putting their hands into the air. Have you ever participated in this action?
- What is the meaning of that action, to you? Why did you choose a particular time and not another to act in this way?

- I often feel vivified and empowered by the music, have you ever felt that way from listening to or participating in it? (If yes) Was this ever a trigger for you acting out?
- I've heard a Pastor at Crossroads mention people telling him that when God speaks to you get goose bumps; do you believe this to be true? (If yes) Have you ever experienced this for yourself? What were you doing when it happened?

Data

Somatic experiences, particularly goose-bumps, were considered to be evidence of religious truth by all four subjects. The emphasis placed on them by the individual, however, varied considerably, as did the circumstances allowable in counting an experience as spiritual. (See: Findings) Abraham explained how so-called goose-bumps could be a sign that God was speaking to him through the Bible. He said, "...when reading God's word, I get some understanding, it often is accompanied by, just kind of a, washing over my body...[of] goose-bumps..." Daniel also described a similar phenomenon. He said, "Sometimes you get goose-bumps... Holy Spirit goose-bumps... it's just this sense of, you feel like, like, a breakthrough, almost, Wow, things get clearer, it's almost like a light inside... something just... resonates... in some ways it's a somatic resonate, I hate to say clicks, but something just seems... more right, than ever before" (See: Findings). Eve and Moses, who also described goose-bump-like feelings, believed them to be the physical body's way of expressing spiritual things. The latter said that goose-bumps was the body reacting to "the presence of the Lord. I don't know how else my body's going to handle it, you know, this physical body's so limited..." Others described physiological reactions of a different nature. For example, the tingling effect of music on the body was considered evidence by several of the subjects, such as Moses and Daniel. Moses was more conclusive in this area than was Daniel, but both agreed that it could be the result of the Holy Spirit. In his interview soon after a particularly energetic service, Moses told me, "Something happened in there today... I think... they [the congregation] just connected with the Lord. You know, [the Bible] says there's cherubim... always going around the throne... And I think in some ways we're seeing a little bit of what they see." Though each individual

showed a preference for certain types of somatic experiences, all counted one or more as a source of religious conviction.

The Bible, that is, the claims presented in the Bible, were also considered to be a source of religious conviction by all four subjects. Three subjects stated that the Bible had been the initial source that convinced them of the truth of Christianity. According to Eve, there was and is “nothing else.” She truly embraced Christianity as a child when she was “watching Billy Graham on TV” and “realized that what he was saying was true.” She acknowledged, however, that she had been raised believing the Bible since birth. Abraham, too, was raised a Christian, and attributed his conversion to hearing and understanding the message of Jesus as presented in the scriptures. The Bible was also his guidebook: when asked if every other experience had to come back to the Bible he answered, “Yes. I want to say that, but honestly, I feel, I really have to say that, because if I divert from that, whatever. You can do whatever you want.” Daniel also considered the Bible to be the integral source of his first conviction. He told me, “The Bible was very strong to my conversion because [a relative of my wife] would [say]... the Bible says Jesus is the way, the truth, and the life and no one comes to the Father except through Him. And, every time we would say something she would quote the Bible, and somehow it did its work... so I started reading the Bible and really taking it into my heart and became convicted by it. So, the Bible is huge.” Moses, when asked of the Bible’s relevance to his faith, answered, “He sent us a letter, that, if you read it... you can really know what He’s thinking... you can know Him through His letter to us, it’s huge, it’s like everything.” Moses was the only individual, however, for which the Bible was a subservient or secondary source of conviction (See: Findings). All four subjects considered the Bible to be a source of conviction and useful in verifying or rejecting other sources.

A third source of conviction was strange happenings or coincidences that seemed too far-fetched to have occurred without the involvement of an outside agency. This, too, was mentioned by all four subjects. Whatever the occurrence (for they varied considerably), each one made the subjects abandon naturalistic explanations for God-directed ones. Daniel told me a story of his childhood where his father made a point of telling him about a very important report on a medical condition. At the time, he saw little importance in such a thing; yet

for the past twenty years he has worked on researching the very same condition. For Daniel, this was proof that “God has led my life... [and] is in control of my life too.” Moses mentioned an occurrence that took place over a shorter time span. He told me, “I had a dream the other night... about a young guy in the church here.” It was “out of the blue” and seemed odd, so he decided to approach the boy in question to tell him of it. After he had done so, the boy reportedly told his father that, “when I was walking over there to see Pastor I knew exactly what he was going to tell me.” The father later related this comment to Moses. This dialectical relationship, where both individuals had a peculiar subject on the mind for no discernible reason and informed the other, was proof of God’s hand; after recounting his tale Moses told me, “A lot of times [God] speaks to me in that way.” Abraham also made reference to such things. Before making a trip to Israel, he had prayed for God to give him a more charity-oriented heart. At some point on the trip, a man approached him to ask for money, and Abraham gave him all the money he had. When he returned home, he received a gift basket from a woman he had never met. Therein was a check for \$1000, and it had been written “on the same day, and perhaps the same point of time, when I’m giving [charity]... [and] God’s over here, moving this woman to give to us.” He conceded that it could be “a coincidence” but that he believed it to be “God speaking, you know, that was a good thing you asked of me.” Eve concurred with the rest on this point. Responses [like Abraham’s] were usually conditional, but still embraced belief. By the reasoning of the four incredulous subjects, a highly improbable event at the very least implied the hand of God.

A fourth source of conviction was positive emotions, which were felt and asserted by all four subjects. This category includes internal head feelings of joy, peace, comfort, and hope. It was also a compelling source: none of the subjects expressed skepticism as to the nature of such emotions (See: Findings). Most of the subjects talked of these as if it were commonplace to understand such emotions in a spiritual way. Daniel discussed all of them with no sign of reservation or doubt, unlike other sources. Eve, too, found them to be proof of the Holy Spirit’s work. Upon being asked of the evidence the Holy Spirit gives her of his presence, she answered conclusively, “He gives me great joy... no matter what the circumstances are around me... He gives me peace.” She then echoed the Bible’s explanation of this phenomenon,

stating, “It’s the fruit of the Spirit, because when the Holy Spirit resides in you, you have love, joy, peace, patience...” Abraham, too, described an “internal peace” and “an inner confidence”; this was the only source of conviction besides the Bible which he spoke of without reservation. Moses expressed a similar opinion, explaining, “The evidence is real simple, Jesus said, I must leave you, so that the Comforter can come, so, first off, I feel comfort, a sense of peace like you said.” Positive emotional feelings were seen as originating from a divine source and were proof of the presence of God.

A fifth source of conviction that was mentioned solely by Daniel was the universe itself and the complexity of living things. Said Daniel, “I really think it takes more faith to be an atheist than it does to believe, based on what I think the scientific evidence points to, you know, the universe had a beginning, and if there’s a beginning there’s a beginner, and... the design in life... and if there’s a design it means there’s an intelligence beyond the design... [T]he most primitive form of cell... they have as much specified complex information as 1000 encyclopedias. And... I don’t believe that can happen by chance.” His argument stems from the absence of a materialistic explanation for certain phenomenon. Akin to the third source of conviction, personal incredulity at an event occurring “by chance” is grounds for a divine explanation. In describing this source, Daniel mentioned several authors who he agreed with or had read, such as Michael Behe, Lee Strobel, and William Lane Craig (these are current figures in the apologetics or Intelligent Design movement). All three use “scientific” arguments to prove the reliability of Christianity. It seems reasonable for Daniel to ascribe to their claims: as an academic professional who described himself as a “wired skeptical,” he wanted more reputable proof for Christianity than the circular reasoning of believing the Bible because the Bible says so (See: Findings).

Findings

Daniel, Abraham, and potentially Eve described experiences that accurately mirror the previously described “realization sensation” (See: Methods). Their episodes began with the sense that they had experienced “a breakthrough”, had a “revelation”, or gained “understanding.” For Abraham this had occurred during reading of the Bible. He explained, “...when [I’m] reading God’s

word [and] I get some understanding, it often is accompanied by, just kind of a, washing over my body... [of] goose-bumps.” Daniel was vague as to the setting for his own experiences but seemed to mention the Bible being involved. Eve spoke of her spirit “Bear[ing] Witness” when someone “say[s] something that’s a revelation to me... something that I was trying to understand.” When she was asked if it was similar to making a connection, she said, “Absolutely, it’s just like, that’s awesome, that’s it, that’s the answer.”

In all three cases something “click[ed]” and two of the three described this as being akin to a light bulb turning on inside. For Daniel and Abraham, the immediate response to their discovery was a washing over sensation of goose-bumps. The two men’s described experience matches the realization sensation on all points. Moreover, Daniel explicitly described the sensation as a “somatic resonance” while Abraham mentioned being “emotionally... pricked.” Both men believed these experiences to be the work of the Holy Spirit, or God himself, and used them as evidence of spiritual things. Eve’s, while containing the mental components, lacked a description of goose-bumps. For this reason, her description only implies the realization sensation and may actually be relating a separate phenomenon. Nonetheless, the experiences of these three provide evidence for the conclusion that Christians, when recounting an experience where God spoke to them through the Bible, may mean or be describing the realization sensation.

While somatic experiences were noted by all individuals as a source of conviction, Moses alone considered such an experience to be his primary evidence of spiritual things and what led him to become a Christian. He stated that, while lying in the snow hoping that God would respond to his supplications, he experienced “...a heat, that overcame me. And it was just amazing; I never experienced anything like it in my life. I just really, just really experienced the presence of God...” Follow up questions were able to elicit further details: Moses explained that it felt like “Something in my body...just went into an over acceleration... my body kicked into another gear... And I just knew that I knew. That there was a God.” When he was asked of his opinion of the Bible before this experience, he responded, “I was aware of it... it was like everything else, it was just a bunch of philosophical thoughts and opinions of people... and, you know, why are their thoughts any better than my thoughts?” The day after his somatic

experience, he woke up "...at 4 o'clock in the morning, and all I could do is just read the Bible, I just wanted to find out who this God is." His statement runs contrary to the experiences of the other three interviewees who claimed the Bible to be their original source of conviction. For Moses, the Bible was a secondary source of religious truth in his conversion that was subservient to a somatic experience.

On a similar note, Moses' and Eve's conviction seems to conflict with parts of an explanation (Luhrmann, et. 2001), which claims that coming to understand God's role in an "unusual, vivid spiritual experience" is a learning experience. According to Luhrmann, identifying God's involvement (or lack of involvement) in a situation requires social instruction and may take multiple tries. Says Luhrmann, "...we work with a theory that learning to experience God depends on *interpretation* (the socially taught and culturally variable cognitive categories that identify the presence of God), *practice* (the subjective and psychological consequences of the specific training specified by the religion: e.g., prayer), and *proclivity* (a talent for and willingness to respond to practice)." In regards to the interview with Moses, two of Luhrmann's categories (proclivity and practice) seem questionable: Moses elucidated no previous cultural or social training that led him to the conclusion that God had caused his somatic experience. On the point of interpretation, conflict appears: Moses had received no "specific training specified by the religion." There was no doubt in his mind that the heat-related somatic experience was God, even though he explicitly stated that he had "never experienced anything like it in [his] whole life." This was something new, something completely alien to his mind and body; yet he still believed it to be God, and immediately so, if his recollection is accurate (See: New Questions).

Eve, too, was willing to make conclusive statements on phenomena that she was unfamiliar with. Towards the end of our interview, she read me a journal entry about something her husband had experienced at church after she finished describing this, the claims were verified with her husband, who was present in the house). During worship time, he reportedly felt "something going into his wrist... it was painful...it was like a spike..." Her husband was apparently unsure as to the meaning, but Eve thought it meant that Jesus was helping him understand part of what he suffered through on the cross. She saw it as "a powerful example of the love that [Jesus] has for

[my husband]." Eve asserted this experience to be from God despite the fact that she had never experienced it for herself. This highlights Luhrmann's emphasis on practice and perhaps proclivity while calling the need for interpretation into question. On the count of practice, Eve was in a situation prescribed by the religion for encountering God ("worship time" at church). On the count of proclivity, she connected this unprecedented experience to the Bible, a source of conviction. However, there is neither Bible verse nor social/cultural teaching, to this researcher's knowledge, that asserts that physical pain is evidence of God's presence. Where exactly did she get this idea? Typically, pain in the wrist would be interpreted as a sign of an internal problem or a strained muscle and would be dealt with medically. Perhaps the context directed Eve's thinking in this situation (See: New Questions). Overall, both Moses' and Eve's statements call into question certain aspects of Luhrmann's explanation.

The responses also suggest that why one converts and what one finds to be compelling proof of religious claims may depend on one's period of life at the time of conversion. Abraham, who was raised Baptist from birth and stayed so throughout his life put primary and solitary emphasis on the Bible. When he "was 5 years old" he had "the simple understanding of... what I believed in, the Bible's teaching." This occurred chiefly at an evangelical service where a preacher used "...the word of God. And the law of God converts the soul, the Scripture says." In discussing other sources of conviction, Abraham often added the precursory statement, "Scripture says," and ensured other experiences lined up with the text. He stated, "It [corporate worship, emotional experiences] needs to be based on the word of God, and not a superficial understanding, you want to go deep." Somatic and emotional experiences were acknowledged, but downplayed; Abraham tended to use conditional phrases when he discussed them, such as, "You know, I believe *at times* [goose-bumps] is [proof of God's presence]" (Italics mine). He also made sure that I understood that somatic experiences "...don't have to be [evidence of spiritual things]. Because my emotional experiences can be completely wrong..." As should be clear, Abraham was raised believing in the Bible and continues to rely on it as his primary source of spiritual truth.

The three other subjects (barring Abraham) converted to their current faith in middle age and placed more emphasis and confidence on emotional

or somatic experiences. They still believed the Bible to be an adequate proof, but spoke more conclusively of other phenomena. Moses, as previously mentioned, considered a somatic experience to be his elementary proof of God's existence. Like Daniel, Eve was raised in a form of Christianity but left that faith entirely for a time: she told me, "My whole family, is believers, and I [was] like... I don't want to hear that anymore." Now that she has returned, she claims the Bible to be compelling truth of Christianity and claims it to be "everything" to her faith; but in answering a question on her walk with God/relationship with Jesus [a fundamental trait of Evangelicals] she exclusively discussed emotional satisfaction. She was also much more prone to emphasize or speak conclusively of somatic experiences as a source of conviction. At the end of our interview, Eve asked me if she could read me a passage from a "powerful little book." She proceeded to do so, and the words/the imagery evoked a tingling sensation in my skin and neck. When I mentioned this sensation to her, she replied, "Yeah, because something's bearing witness in you, in your spirit. It's bearing witness about the truth." Eve used no conditional phrases here, which were the rule with Abraham. These examples show that individuals who converted later in life still believed in the Bible but were more willing to unconditionally accept other sources of conviction.

Daniel's responses reinforce this notion that one's reasons for conversion may be based on the period in which one converted. During the interview, Daniel actually stated that there were "two parts to [his] life. There's being raised Catholic, and falling away, and then coming back a follower of Christ." He also stated that he has held "different levels of belief." When he was younger, said Daniel, "I... had sort of a child like belief" but now through the works of apologists and scientists he has "take[n] it a notch higher... [and] come to the conclusion that [his] faith is a reasonable one." A follow up question was asked, that being, "Would you say that when you were a child it was because of tradition or upbringing?" He responded, "Yeah, very much. It was, very much." These comments seem to imply that a person may have to re-evaluate and lay new foundations for their faith at different points in their life. That is, what a person finds to be compelling proof of their faith at an early point may no longer convince them in subsequent years. As a child, Daniel believed because he had been taught to; now as an adult he believes because of his own study and efforts.

Based on Daniel' and Abraham's comments, there seems to be a correlation between the reasons for one's conviction and the life period of the individual at the time.

The interviews with Daniel, Eve, and Moses also tend to show that individuals who embraced a religion later in life did so in times of crisis. In the case of Daniel and Moses, these crises involved the realization of human mortality. In the former, this came by the death of family members, and in the latter, by a diagnosis of cancer. Daniel explained that "...we converted shortly after my [family member's] death... And it became very obvious to me that, it shook the foundations of my life [which was] I must be the power." A similar experience occurred with Moses, who expounded how years prior when he was diagnosed with cancer he "began to search for God..." Eve was vague about a crisis, but she mentioned being wounded emotionally and having family related problems. After having found a place at Crossroads, she reports, "My spirit... it's been completely healed." All three of these episodes involved instability, were emotionally detrimental to the person, and occurred within close proximity to that individual's conversion (See: New Questions).

Daniel's interview supports the possibility that what one considers to be a compelling source of religious conviction (and why) can be influenced by education or occupation. In discussing the Bible, the other three subjects used no extraneous evidence as a base for their conviction; a solitary reference was made by Eve to the Bible being trustworthy because prophecies made in the Old Testament came true in the life of Jesus. Such was not the case with Daniel: he backed up his belief in the Bible by making reference to empirical evidence which collaborated the Bible's account. Said Daniel, "I have studied... a lot of the archaeology stuff and my beliefs about the Bible... the Hittites, some people say, the Hittites never existed, and then they dug up the Hittite library. The Bible seems to stand the test of time, more than any other ancient book." Another example was given concerning textual reliability: reportedly, early Christian letters quote almost the entirety of the New Testament. Thus, one does not have to assume the existence of an unadulterated text, but can verify this using contemporary sources. These proofs have a more empirical basis than others provided by him and other subjects. As an academic professional who publishes research, it makes sense for Daniel to place trust in such proofs, as they align more accurately with his

thought process. Daniel's experience implies that individuals can be shaped by both the type of the work they do and the education they have received.

Sources of conviction that can be found explicitly described in the Bible were embraced more intensely than those that seem to lack textual support. For example, all four individuals spoke with certainty that emotions such as peace, joy, and confidence were from the Holy Spirit (See: Findings). These emotions and more are explicitly described in the canon: Galatians 5:22-23 gives a list of moods that may be considered "Fruit of the [Holy] Spirit" (which includes peace and joy), while John 14:26 describes how Jesus would send "the Comforter" to his disciples. The interviewees, as previously described (See: Findings) placed great confidence in the Bible; some (Moses, in certain situations, and Abraham in all) even used it to discern the nature of external phenomena. It can be safely concluded, then, that their confidence in positive emotion as a source of religious convictions originates from the Biblical writing on the subject. This would also assist in understanding why interviewees were less confident (in general) concerning other sources of conviction. While all the subjects believed that somatic experiences could be proof of spiritual things, they never made blanket statements and considered the context before making such a judgment. Conditional phrases such as "could be" or "maybe" were used, with rare exceptions (such as Moses' heat related somatic experience). Clearly, the Bible is still an important source of religious conviction to Christians and may set standards for gaining belief from other things.

Limitations:

The responses of those interviewed, for the great majority, indicate the validity of this study and the conclusions drawn from it. Interviewees seemed to have a good grasp on the concept being pursued as well as what exactly was meant by somatic experiences. When they were unsure, they asked questions. For example, when Daniel was told that somatic experiences were being researched, he asked, "What kind of experiences?" for clarification. After being informed of what was included in that category, he immediately began to talk about his experiences with goose-bumps. The only potential problem encountered was the occasional difficulty of separating emotions from somatic experiences when subject's talked of both in close vicinity. This interchangeability is

understandable as both emotions and somatic experiences involve sensations within the head. Both are also experienced as reactions to the outside world. For example, Abraham mentioned being "emotionally" pricked by finding understanding within the Bible, but then immediately went on to describing goose-bumps. To him, goose-bumps and "head feelings" were counted as emotions. The line between the two can be hard to draw, thanks to the similar characteristics of both. However, particular terms could be used to divide these apart (goose-bumps, tingles, etc. meant somatic experiences, specific named moods are emotions). Thus, this was no significant problem and could be easily sorted out. Overall, the subjects seemed to grasp the concepts being studied and answered questions accurately.

How far this study can be generalized is uncertain and probably limited as the group interviewed was considerably limited both in number and variety. It encompassed a total of four individuals, all of whom were Protestant and Evangelical. This sample was also unrepresentative of the population of the area and the church: all of the subjects (one female and three males) were white, middle class, and middle aged. Moreover, three of the subjects had received at least some level of education beyond the high school level (and for two this education was extensive.) The background of the subjects, though, provides some assistance in extending the scope of this study. Of the four individuals, only two were raised in the same faith. Moses came from a non-religious background, as was previously mentioned, which is helpful in that he was less influenced by religious believers. Daniel was raised Catholic, but left his faith during his college years. He experimented with various Eastern religions for a time before coming back to a different form of Christianity. Both Abraham and Eve were raised Baptist, and at least Abraham has stayed very much so. This variety of past faiths lends credibility to the idea that an emphasis on somatic experiences may not be a merely Protestant Evangelic phenomenon, but rather a human one. Overall, it would still be necessary to have a more diverse sample in order to reveal some general principle that would help in explaining the source of religious conviction.

The reflexivity of subjects seems to be limited if present at all. Leading questions that made explicit reference to the details of particular somatic experiences were purposefully left till the end; this ensured that individuals would not highlight such experiences unless they felt that they

were crucial to their faith. There were also times when answers that would have assisted the thesis of this research were not given under further inquiry. Toward the end of Eve's interview, she mentioned a potential connection between speaking in tongues (to her, a Holy Spirit-given ability) and goose-bumps. A question was asked as to which came first in the sequence, and she indicated that glossolalia did. By that point, she was aware of the focus on the primacy of somatic experiences, but did not give a response conducive to that conclusion. Furthermore, when inaccurate conclusions were drawn, the subject (always, as far as could be discerned) spoke up to correct the mistake. For example, based on Moses' claim that "the Bible is God's letter to Christians," a statement was made to Daniel as part of a question which assumed this to be his view as well. He corrected this assumption immediately, and started his response to the full question by saying, "I don't think that the Bible is a letter from God to Christians, I think that the Bible is a letter from God to man..." For these reasons, it can be safely concluded that the presence of the researcher had little effect on those interviewed.

Suggestions for future research:

Further research could be conducted to determine how influential somatic experiences are overall in creating religious conviction. This researcher feels that the main deficiency in this study was the lack of a diversity of religious faiths. An interesting group to focus on in the future may be Muslims. Recently, this researcher heard the Qur'an being sung/recited by an imam, and this elicited a somatic response. Moreover, the realization sensation seems to be best evoked when a connection is made between two ideas. The Qur'an is a vast piece of literature with sections divided seemingly haphazardly; finding new details on existing characters in separate Suras might be enough to evoke goose-bumps. It would be interesting to see if Muslims likewise experienced such sensations and gained religious conviction from them. If this is the case, it is possible that the same thing could be true of Jews who recite or study the Torah. Ensuing research could also deal merely with other Christian sects: if it is a wholly Christian phenomenon, perhaps one particular sect is responsible for the spread of this idea. The Gregorian chants of Catholic monks could presumably incite such responses, as could perhaps particularly vivifying organ music. Subsequent

research, to remain free of bias, should also target those groups that are hostile or silent concerning somatic experiences, such as fundamentalist Baptists. It has been shown by comments in this study that somatic experiences can still be encountered in a non-music based setting. Thus, even those that frown on emotional music but have individual's reading the holy text could potentially have somatic experiences. If somatic experiences were encountered, would predisposed individuals still gain conviction from them? Investigating the opinions of both those in an environment that is and is not favorable to interpreting somatic experiences in such a way would be the best route to a general answer.

Another study using the same religious group or Crossroads itself would still be beneficial and could be done with a more diverse sample. A more balanced study could make use of several elderly and young-adult subjects; this would be more representative of the population as a whole as well as most churches. Likewise, future subjects could be selected in ways to better avoid bias. During Crossroads' services, one can tell (with relative certainty) which individuals are receptive to certain somatic experiences and which are not (for example, people can be observed throwing their hands up when the music evokes a tingling sensation). A point was made in this study to avoid bias by not purposefully picking those who seemed the most responsive to such things, but this was hard to accomplish. Preliminary questions, or more extensive observation, could sort out a future group of those who do not seem inclined towards somatic experiences. It would be interesting to see if this group would still consider somatic experiences to be a source of religious conviction. Such a study would show whether or not the four individuals researched are typical within the group or merely a radical minority. Further research along this line would help reveal the influence and scope of somatic experiences for neo-charismatic Christians.

It would also be beneficial to discover more details concerning what exactly is required and involved in the creation of a somatic experience. They clearly involve the nervous system, in at least some capacity, but what is it about connecting two ideas or hearing particularly powerful music that evokes such feelings? Does the realization sensation come on from the connection of synapses in the brain (which seems to be where the sensation starts) or is it merely a worked up mind releasing stress? As music is involved, could certain specific keys or chords bring on tingling sensations by

making certain neurons fire in a peculiar pattern? Most religions and rituals, regardless of their locality, make use of music in some capacity. If somatic experiences are a widespread source of religious conviction, this could give a potential explanation for the cultural retention of certain rituals (See: New Questions). Somatic experiences as a whole could be studied on a psychological or biological level by attempting to pinpoint integral regions of the brain that are involved in the reaction (it is possible that literature already exists that provides such an answer). Further sociological study could be targeted on those individuals who have experienced these sensations. In-depth and specific questions could then be asked concerning the person's mood at the time, what they were thinking of, where they felt the sensation begin and end, etc. Whatever the branch of science, any of these approaches would help light the darkness surrounding these unusual sensations and their origin.

Based on Moses' conversion, further research could be conducted to see if other individuals have embraced a religion thanks to a somatic experience. This would not need to be limited to actual interviews with believers, though this would be helpful. Instead, or in cooperation, historical research could be conducted on the autobiographies of past and current believers. This medium may be a great source of information based on a recent finding. A few weeks ago, this researcher read through a chapter of *Godless* (which chronicles the conversion and de-conversion of Dan Barker, an evangelical pastor turned atheist). According to the author, he was raised a Christian, but gained his real conviction and direction one evening at a "revival meeting" when he believed that he was "communicating with God." Says Barker, "[That evening] I experienced strong inner sensations that I could only describe as "spiritual"... I had never had these feelings in any other context, and since the "spirit-filled" environment triggered them, I knew that I had confirmation of the reality of God." Such sensations, which seem to be brought on by the music, have been previously described as a source of conviction by the four individuals in this study (See: Findings). Their remarks and Mr. Barkers' remarks assist in the conclusion that other neo-charismatics may gain conviction from such experiences. As the neo-charismatic camp seems the most promising at this time, its literature could be dredged to find the sources of conviction leaders and followers embraced or embrace.

Further study could try to ascertain exactly why certain devotees believe somatic experiences to be a source of religious conviction. To this researcher's knowledge, neither Bible verse nor Christian dogma supports the idea that such feelings are proof of the Holy Spirit's presence. Why, then, do these Christians believe that they are? Either the Bible is not the filter for all other experiences, and Christians merely play lip service to it at times, or a description of certain somatic experiences can be found in the text. There are enigmatic references in the Book of Ephesians and the Book of John that may potentially be referring to somatic experiences. For example, John 16: 13-15 reads in part, "When the spirit of truth comes he will guide you in all truth... He will glorify me, for he will take what is mine and will declare it to you." In Ephesians 1:17, the Holy Spirit is called "the spirit of wisdom and truth," while Ephesians 5: 18-19 mentions that one can be "filled with the Spirit" by "psalms, hymns, and songs." Could the declaration mentioned in the Book of John involve the realization sensation, which accompanies understanding and the connecting of ideas? Likewise, could this filling of the spirit spoken of by the Book of Ephesians refer to the tingling sensations brought on by particularly powerful music? Ephesians 3:16 may reinforce this conclusion: it says that God will "strengthen [believers] with power through his Spirit in your inner being." Even with these verses, neither of these interpretations is conclusive; nor is any of this evidence beyond the circumstantial. The best way to verify such an explanation would be to search through the writings of early church fathers to see if they interpreted such passages in a similar way. Or, interviews could be conducted where Christians would be asked to interpret these passages as they understand them. Until either of these can take place, a reason for the acceptance of somatic experiences as evidence of spiritual things is uncertain.

The context surrounding a somatic experience could be explored further as it seems to have importance in the interpretation of such as a source of religious conviction. Humans who are unfamiliar with these experiences and then have them during a ritual may see that ritual as endowed with some holy fire. This explanation was explicitly given by Dan Barker earlier in this section for his belief, and this may be the case with others. Something to this effect was mentioned by Abraham, who gained faith from somatic experiences even though he noted "Not [because] the Bible tells me that

[they're spiritual].” Instead, he implied that he believed them to be from “the spirit of God” because “when reading God’s word, [and he] gets some understanding, it often is accompanied by... goose-bumps...” That is, because the realization sensation frequently occurred within the context of a religiously prescribed ceremony, Abraham believed it to be proof of God’s presence. Based on these items, a study could focus on individuals who have had somatic experiences (preferably, in both secular and religious settings). Likewise, it would be best to include both those embrace a religious faith and those who do not. Each person could be asked of their experiences, under what circumstances they happened, their interpretation of a few, and why they came to such a conclusion. This researcher would expect a connection between those who had somatic experiences under religious circumstances and those who interpreted them as a source of religious conviction. Somatic experiences under a secular context should not be expected to be viewed as commonly as a source of religious conviction. Interviews of this nature would help elucidate whether or not a particular setting is necessary for a somatic experience to be interpreted as “spiritual.”

New questions

The experience of several subjects who converted later in life raises a question in regards to the origin of their desire to pursue the divine. Daniel and Eve were raised in various branches of Christianity, which they later rejected; they returned to a different form of Christianity later on in their lives soon after a crisis. It can be safely concluded that the idea of an afterlife and a loving God had already been implanted in their mind at a young age. This, most likely, provided an impetus for their return. The same conclusion cannot be as easily made for Moses, who also converted in a crisis: he was not raised religious and had little if any previous experience with Christianity. Moses did describe an employee at his workplace that may have influenced him: this individual “...believed that [the Bible] was true, that everything in it was true.” Later in the interview, Moses made the comment that this person was unaware of his search for God, which would seem to indicate that discussion between them on the subject was limited, if it did occur. It is possible, therefore, but not likely, that Moses was socialized into accepting the existence of the Divine before beginning his search. If this is the case, then his yearning for God comes as no surprise, as the idea of pursuing God

had already been placed in his mind by his co-worker. If this was not the case, then an obvious question is: from whence did his desire to pursue God arise? Is a yearning for some form of truth greater than oneself a part of the natural human being, or is it a socially constructed urge? May it be that, in crisis, the human mind simply latches on to whatever possibility seems most immediate? As it is impossible to completely isolate an individual from religious influence, an answer to this question is difficult to determine. Though there are possible intrusions of influence, the source of Moses’ desire to pursue religion is unclear.

The fact that interviewees considered somatic experiences to be a source of religious conviction raises the question of whether or not certain rituals are purposely used because they can elicit such responses. Whether or not current rituals were designed this way would be impossible to ascertain, if they were designed at all, as many have their root in antiquarian practices. For example, a modern practice known as the altar call can be commonly found in Baptist churches. This ritual involves the offering of salvation to unbelievers and is usually accompanied by the playing of particularly inspiring music. Of course, the use of music in religious rituals is hardly original; it can be inferred from certain shamanistic cave paintings in France that this practice may date back to nearly 30,000BCE. Thus, if there was a single motive behind the “choice” to use music in a religious context, it cannot be known. However, individuals in the religions today still choose to use music in certain ceremonies. Do pastors or choir directors consciously choose a piece because of its ability to (to use an Evangelical term) “make the Spirit move”? Or are they simply following the traditions of their church with little interest in the results? Do imams recite the Qur'an in such a melodious fashion to send tingles across the skin of their listeners and endow the performance with a spiritual energy? Or were they merely trained to do this as a child and have received no other explanation as to why this must occur? This should not be read as an accusation that religious leaders are manipulating their followers’ faith. If in fact a conscious decision is being made to use rituals because they incite somatic experiences, those making this choice may honestly believe that these rituals are necessary to feel the presence of the divine. At this current time, it cannot be concluded one way or another whether somatic experiences are merely the unintended consequences of a ritual or whether they are the intended result.