

The Journal of the
**Utah Academy of
Sciences, Arts, & Letters**
Volume 88 • 2011

**Includes selected and refereed articles from the
2011 Annual Conference
held at
Salt Lake Community College
April 8, 2011**

Editor
Kristin L. Kraus

Copyright 2012, Utah Academy of Sciences, Arts, & Letters.

Copyright reverts to authors after publication. Printed in the United States of America by Brigham Young University Academic Publishing, Provo, Utah. Neither the editors nor the sponsoring organization assumes responsibility for statements of fact or opinion made by the contributors.

ISBN-13: 978-0-9815501-5-2

All Rights Reserved

Board of Editors

Arts: **Marcus A. Vincent**, Utah Valley University
Biological Science: **Erin O'Brien**, Dixie State College
Business: **Jonathan H. Westover**, Utah Valley University
Education: **David D. Williams**, Brigham Young University
Engineering: **Doran Baker**, Utah State University
Exercise Science and Outdoor Recreation: **Shaunna McGhie**,
Utah Valley University
Letters—Humanities, Philosophy, Foreign Language:
Jim Gustafson, Southern Utah University
Letters—Language and Literature: **David Allred**, Snow College
Physical Sciences: **Jean-Francois Van Huele**, Brigham Young
University
Social Science: **Kristine Olson**, Dixie State College

2011-2012 Utah Academy of Sciences, Arts, & Letters Officers

President: **Nichole Ortega**, Utah Valley University
President-Elect: **Erin O'Brien**, Dixie State College
Past President: **Spencer Blake**, Salt Lake Community College
General Secretary: **Jennifer Schultz**, Utah State University
Membership/Newsletter: **CoCo James**, Salt Lake Community
College
Treasurer: **Karl Israelsen**
Scholarly Communications: **Peter Kraus**, University of Utah
Member at Large: **Dwight Israelsen**, Utah State University
Webmaster: **R. Steven Turley**, Brigham Young University

Cover Photo: David C. Ripplinger, Brian E. Anderson, and Timothy W. Leishman. “Instantaneous frequency tracking of partials of a string vibrating at large amplitudes using a high-speed camera.”

Utah Academy of Sciences, Arts, and Letters

History: Founded 3 April 1908, the Utah Academy of Sciences was organized "to promote investigations and diffuse knowledge in all areas of science." Beginning in 1923, the Academy started publishing the papers presented in its annual meetings in *Proceedings*. In June 1933 at the annual meeting, the Academy was enlarged to include arts and letters, and the name was changed to the Utah Academy of Sciences, Arts, and Letters. Articles of incorporation and non-profit organization status were accepted by the Academy membership at the spring meeting in April 1959. In 1977, the name of the journal of the Academy was changed from *Proceedings* to *Encyclia*. It became a refereed journal at this time. In the mid 1980s, the scope of the Academy was expanded further to include (1) business, (2) education, (3) engineering, (4) library information and instruction, and (5) health, physical education, and recreation. Beginning with the 1998 issue, the journal became *The Journal of the Utah Academy of Sciences, Arts, and Letters*.

Annual Meeting: The Academy's annual meetings are normally held in the spring on one of the Utah campuses of higher education. The plenary session is called the Tanner Lecture, endowed by Mr. O.C. Tanner in 1986.

Best Paper Awards: The best paper presented in every division is given a cash award, which is presented at the Academy's "Awards Evening" held the following fall.

Distinguished Service Awards: The Academy recognizes outstanding contributions to teaching and scholarship by means of annual Distinguished Service Awards, alternating every other year between disciplines.

Membership: When the Academy was founded in 1908, membership was by nomination, ratified by the Council, and elected by a "three-fourths votes of members present." Today, the Academy's membership is available by application.

Institutional Members: All Utah institutions of higher education are members of the Utah Academy. The Academy appreciates their patronage.

Publication Policy

The *Journal of the Utah Academy of Sciences, Arts, and Letters* publishes works in all of the fields of study encompassed in the Academy's mission. Papers published in *The Journal of the Utah Academy of Sciences, Arts, and Letters* are drawn from papers presented by members in good standing at the annual conference of the Utah Academy. To qualify for publication, the papers must be recommended through a refereeing system.

Presenters are encouraged to publish their paper in *The Journal of the Utah Academy*. *The Journal's* criteria are that a submission is (1) fresh, meaningful scholarly insight on its subject; (2) readable and well written; and (3) of general interest for an academic readership beyond the author's field.

If you wish your paper to be considered for publication in *The Journal*, please submit a Microsoft Word document to the section editor of the appropriate section by the indicated deadline. Contact information for the section editors is available on the Utah Academy's website (www.utahacademy.org).

The Journal of the Utah Academy is a refereed journal. Editorial responses will be forthcoming after the resumption of school the following fall when referees have returned their comments to the division chairs.

Papers should be between ten and twenty double-spaced pages. Detailed instructions to authors are available at http://www.utahacademy.org/Instructions_for_Authors.pdf.

Among the bibliographic services listing at Bowker Serials Bibliographies and The Standard Periodical Direction. Indexing and abstracting services that cite articles in the journal include Arts and Humanities Citation Index, Biosciences Information Services, Current Geographical Publication, Chemical Abstracts, Mathematical Reviews, MLA Biography, Sociological Abstracts, Excerpta Botanica, Social Planning, Policy and Development Abstracts, Language and Language Behavior Abstracts, Index to Scientific Technical Proceedings, and Index to Social Sciences, and Humanities Proceedings.

The Journal of the
**Utah Academy of
Sciences, Arts, & Letters—2011**

AWARDS

Distinguished Service Award	13
Academy Fellow 2011	14
John and Olga Gardner Prize	16
Tanner Lecture	17
Awards Evening Lecture	18
Honorary Member	19
2011 Best Paper Awards	20

ARTICLES

ARTS

<i>Night Journey: Insight into the Subconscious of Martha Graham</i>	23
Lauren Harris Brown, <i>Utah Valley University</i>	

BIOLOGY

<i>Batrachochytrium dendrobatidis</i> Fungus Update for <i>Hyla arenicolor</i> in Zion National Park	33
Britnee Eng and Curt Walker, <i>Dixie State College</i>	

BUSINESS

- “Right to Work”: Has It Become the “Right to Fire”?**
An Editorial Review 43
Troy Callahan, *Utah Valley University*
- Teaching Theory Versus Practical Use: The Case of the Modern Portfolio Theory** 52
Leo H. Chan, *Utah Valley University and Tan Tao University*
- An Analytical View of China as a Software Outsourcing Outlet** 64
Taowen Le, *Weber State University*, Wayne Huang, *Ohio University*, and Jin Zhang, *University of Wisconsin Milwaukee*
- The Impact of Workplace Wellness Programs on Decreasing Employee Obesity and Increasing Overall Health** 91
Skyler Charles MacDonald and Jonathan Westover, *Utah Valley University*
- The Challenge: Experiential Education in Theory and Practice** 114
Peter Robinson, Laurent Josien, and Rachel McGovern, *Utah Valley University*

EDUCATION

- Practical Theory: A Teacher’s Guide to Making Research Writing Theory Useful for Students** 136
Angie McKinnon Carter, *Utah Valley University*
- Developing Instruction for the Diverse Classroom: Collegial Collaboration in Action** 154
Vessela Ilieva, *Utah Valley University*
- “Quite Probable That They Will All Pass”: Teacher Training at Snow College** 172
David Rosier, *Snow College*

Time for Science? A Study on the Use of Instructional Time for Teaching Science at the Elementary Level	186
Mary Sowder and Stan Harward, <i>Utah Valley University</i>	

ENGINEERING

Side-force Amplification on an Aerodynamically Thrust Vectored Aerospike Nozzle	204
Shannon D. Eilers, Matthew D. Wilson, Stephen A. Whitmore, and Zachary W. Peterson, <i>Utah State University</i>	

Closed-Loop Attitude Control Using Fluid Dynamic Vectoring on an Aerospike Nozzle	230
Nathan Erni and Crystal Frazier, <i>Utah State University</i>	

BYU's Entry in the University Rover Challenge: Making a Simulation Mars Rover	270
Alexander R. Jafek, Andrew K. Norgant, and Justin S. Smith, <i>Brigham Young University</i>	

Document Flash Thermography	295
Cory A. Larsen and Gene A. Ware, <i>Utah State University</i>	

Surface Geometry and Heat Flux Effect on Thin Wire Nucleate Pool Boiling of Subcooled Water in Microgravity	326
Troy Munro, Andrew Fassmann, and Heng Ban, <i>Utah State University</i>	

HEALTH, PHYSICAL EDUCATION, AND OUTDOOR RECREATION

A Comparative Analysis of Prioritizing Professional Behavior between Undergraduate Students of Exercise Science at Utah Valley University and Future Employers	343
Bret H. Boyer, Paige Denney, M. Bohne, and Jason Slack, <i>Utah Valley University</i>	

Eat to Live, Live to Dance: Preventing Eating Disorders in Dancers	348
Kristy Jo Hunt, <i>Dixie State College</i>	

Motivational Behaviors of Elite Senior Athletes 368

M. Vinson Miner and Bret H. Boyer, *Utah Valley University*

LETTERS—FOREIGN LANGUAGE, HUMANITIES, AND PHILOSOPHY

Oedipus's Family in José Luis Borau's *Furtivos* and Iciar Bollain's *Flores de otro mundo* 373

Matías Martínez Abeijón, *Southern Utah University*

LETTERS—LITERATURE

Politics of Display: The Museum, the Academy, and the Framing of American Indian Texts 385

Kim Lawson, *Utah Valley University*

Jules Verne's *The Castle of the Carpathians*: Gothic Castles and Science Fiction in a Literary Landscape 401

Ace G. Pilkington, *Dixie State College*

Nostalgic Renderings of Victorian Clothing in John Fowles's *The French Lieutenant's Woman*, A.S. Byatt's *Possession*, and Neal Stephenson's *The Diamond Age* 416

Amanda Scott, *Dixie State College*

PHYSICAL SCIENCES

Instantaneous frequency tracking of partials of a string vibrating at large amplitudes using a high-speed camera 428

David C. Ripplinger, Brian E. Anderson, and Timothy W. Leishman, *Brigham Young University*

Fabrication and Testing of a Strain-Based Carbon Nanotube Magnetometer Structure* 444

Jonathon A. Brame, Johnathan E. Goodsell, *Brigham Young University*, Stephanie A. Getty, Y. Zheng, *NASA*, and David D. Allred, *Brigham Young University*

**Parallel Computer Simulations of Dynamical Relaxation
Processes and Clusters of Galaxies** 455

Brandon K. Wiggins, Anthony R. Thomas, and Brent A.
Sorensen, *Southern Utah University*

SOCIAL SCIENCES

**Generation Noner: A Look at Religiously Unaffiliated
Attitudes at Salt Lake Community College** 472

Karine Agajanian, University of Utah, and Spencer H. Blake,
Salt Lake Community College

**A Second Soul: Exploring Contexts of Language
Acquisition and Personality** 486

Aaron K. Combs, *Dixie State College*

Learning About Peace and Justice Work 495

Josh Gold, *Salt Lake Community College*

The Regressives: Unraveling the Progressive Movement 513

John Howell, *Southern Utah University*

**Determinants of Racial Concentration in Mountain States
Counties, 2000** 526

L. Dwight Israelsen, *Utah State University*, and Ryan D.
Israelsen, *Indiana University*

**Antigay Behaviors Among Young Adults: Prevalence,
Patterns, and Motivators in a Noncriminal Population
(Replicated)** 546

CoCo James, Chad Morrow, and Spencer Blake, *Salt Lake
Community College*

ABSTRACTS 561

DISTINGUISHED SERVICE AWARD

Dr. Clifton Sanders

Salt Lake Community College

Clifton Sanders is the Dean of the School of Science, Mathematics and Engineering at Salt Lake Community College. He received his Ph.D. in Organic Chemistry from the University of Utah in 1990 and a certificate in Biblical Languages from Salt Lake Theological Seminary in 1995. He has worked as a research technician and senior scientist for several Utah biomedical, industrial, and contract research companies. His academic and industrial research has been sponsored by the National Science Foundation, the National Institutes of Health, NASA, and the Department of Defense. He has coauthored several publications, presentations, technical reports, and patents in mechanistic organic chemistry, mass spectrometry, polymer synthesis, and biomaterials technology. He has worked at Salt Lake Community College since 1993 as a Chemistry Instructor, Assistant Professor, and academic administrator. He was awarded the 1995-96 Salt Lake Community College Distinguished Faculty Lecturer Award for community outreach and science education, and he received a Teaching Excellence Award in 1997 from the Salt Lake Community College Foundation. As an adjunct Instructor in Philosophical Theology at Salt Lake Theological Seminary, Clifton developed and taught courses in Biblical Interpretation, Theology and the Arts, and Theology and Science. He has given theology presentations at the Sunstone Symposium, Salt Lake Theological Seminary and Salt Lake Community College. Clifton plays jazz saxophone with the G. Brown Quintet, a local professional group. He provided solo saxophone improvisations for Emma Lou Thayne's reading of her suite of poems "How Much for the Earth" at the March 2008 Nuclear Weapons Symposium at Utah Valley University. He also coauthored (with Michael Minch) the paper "Democracy as Music, Music as Democracy," which was published in *Radical Philosophy Review*, volume 12 (2009), "Art, Praxis and Social Transformation—Radical Dreams and Visions." He is married to Sandra Watkins and they have a 14-year-old son, Nathaniel.

ACADEMY FELLOW 2011

Kathleen Debenham

Utah Valley University

Kathleen Debenham, M.A., C.L.M.A., is currently Associate Vice President for Academic Programs at Utah Valley University (UVU). She is Professor of Dance and served as the founding chair of the UVU Dance Department, as Modern Dance Program Director, and as the founding Dean (interim) for the UVU School of the Arts. Kathie holds an M.A. from Brigham Young University (BYU), where she taught from 1976 to 1994. While at BYU, she created the Dancers' Company and co-directed the Children's and Teen's Creative Dance program and The Young Dancemakers. She received her Graduate Certificate in Integrated Movement Studies and certification in Laban/Bartenieff Movement Analysis (C.L.M.A) from the University of Utah in 1997 and has integrated the Laban theoretical work into her choreography, writing, performance, administrative duties, and life work. From 1985 to 2000, Kathie co-directed, performed, and choreographed for Contemporary DANCEWORKS, a professional modern dance company based in Utah Valley. She has been a master teacher/artist for the Utah Arts Council and taught numerous residencies throughout the state of Utah and beyond. She served as co-chair of the Utah Academy for Science Arts & Letters Arts section from 2001 to 2005. She is a founding member of the Orem Arts Council and enjoys working to increase access to and quality of arts in the community. Kathie has presented the Laban/Bartenieff work widely at regional, national, and international conferences in dance and the humanities and has published numerous articles on somatic approaches to teaching and learning. She and her husband Pat have enjoyed many years of artistic collaboration, with family-making at the heart of their dance-making.

ACADEMY FELLOW 2011

Pat Debenham

Brigham Young University

Pat Debenham (MA–UCLA, CMA) has been an Artist/Educator for more than 30 years. He is a Professor of Contemporary Dance and Music Theatre at Brigham Young University (BYU), is a Certified Laban/Bartenieff Movement Analyst, was a master teacher for the Utah Arts Council for over 30 years, and was co-director of Contemporary DanceWorks for 15 years. He is currently the director of Contemporary Dance Theatre at BYU and has choreographed more than 20 musicals. Regionally, nationally, and internationally he has been an active participant and presenter in dance, theater, and humanities organizations—NDA, ATHE, NAHE, LIMS, CORD, daCi, and NDEO. In addition to workshops and choreography that have been presented internationally, he has published on subjects as diverse as pedagogy, somatics, spirituality, history, and choreography in *Research in Dance Education*, *NAHE Interdisciplinary Journal*, *Contact Quarterly*, *Journal of Dance Education*, and most recently in *Dancing Dialogues: Conversations across cultures, artforms and practices*. Present passions are his three lively daughters, three sons-in-law, 7 grandchildren, his multi-faceted wife, and a desire for good food, thoughtful reading, and engaging conversation.

JOHN AND OLGA GARDNER PRIZE

H. Scott Hinton

Utah Valley University

H. Scott Hinton was born in Salt Lake City in 1951. He received a B.S.E.E. in 1981 at Brigham Young University and an M.S.E.E. at Purdue University in 1982.

In 1981, he joined AT&T Bell Laboratories in Naperville, IL as a Member of the Technical Staff. He was promoted to supervisor of the Photonic Switching Technologies group in 1985 and then Head of the Photonic Switching Department in 1989.

From 1992 to 1994, he was the BNR-NT/NSERC Chair in Photonic Systems at McGill University and from 1994 to 1999 he was the Hudson Moore Jr. Professor of Engineering at the University of Colorado at Boulder, and finally from 1999 to 2002 he was the Dean E. Ackers Distinguished Professor and the Chairman of the University of Kansas Electrical Engineering and Computer Science Department.

In June 2002, he accepted the position as the Dean of the College of Engineering at Utah State University. He has been very active in the scientific community, where he has published over 35 journal articles and 85 conference papers. He has also been active in service to the professional community by serving in leadership positions for numerous technical conferences and workshops.

Dean Hinton has also been awarded 12 patents. His current research is focused on developing systems applications of smart pixels and free-space optical interconnection, synthetic biophotonic systems, and in developing and understanding technology-enhanced learning environments. He was the 2004-2005 President of the IEEE Laser and Electro-Optics Society (LEOS), an IEEE-LEOS Distinguished Lecturer for 1993-94 and is a fellow of both the IEEE and OSA.

TANNER LECTURE

Action Learning: Using our Research and Teaching as a Platform for Change

Warner Woodworth

Brigham Young University

Warner Woodworth is a social entrepreneur and Professor of Organizational Strategy and Leadership at the Marriott School, Brigham Young University, where he teaches OD, Third World Development, Micro-credit, and Social Entrepreneurship. Author of 10 books and over 200 articles and conference papers, he has been engaged in empowering the poor for three decades. He has helped found and/or served on the boards of numerous non-governmental organizations (NGOs), including Enterprise Mentors International (5 countries), Ouelessebougou-Utah Alliance (Mali), Unitus (11 nations), and HELP International (8 countries), among 24 others. In 2010 alone, the NGOs and microfinance institutions he helped launch from his campus courses in past years grew to over 7.2 million clients, raised some \$36 million, and trained over 520,000 microentrepreneurs. He has been honored with the Faculty Pioneer Award for his global impacts from the Aspen Institute in New York City, a Social Entrepreneurship Teaching Award at the World Forum, Oxford University, and the first Peter Drucker Visiting Scholar at the Drucker School, Claremont University in Los Angeles, among other recognitions. He was also appointed by Nobel Peace Prize Laureate Muhammad Yunus to the Advisory Board of Grameen America in New York City.

AWARDS EVENING LECTURE

Culture Wars in the Classroom: Bridging the Ideological Divide with Comparative History

Joel Lewis

Dixie State College

Dr. Lewis earned his Ph.D. in Transnational and Comparative History at the University of Strathclyde in Glasgow, Scotland. His research specialties are Nazi Germany, political propaganda, American youth movements, and the Spanish Civil War. In 2006, he published a book on international youth politics in the inter-war era entitled "Youth Against Fascism." Prior to teaching at Dixie State College, Dr. Lewis taught American, European, and world history at Central Michigan University and Saginaw Valley State University. In addition to his instructional duties, Dr. Lewis has more than 10 years of community advocacy experience in combating violence against women, children, families, and LGBT youth. He currently serves as the faculty advisor for Dixie State College's Take Back the Night campaign against domestic violence and sexual assault.

HONORARY MEMBER

Melia Tourangeau

Utah Symphony | Utah Opera

Melia P. Tourangeau was appointed President and Chief Executive Officer of the Utah Symphony and Utah Opera (USUO) in February 2008. Prior to joining the USUO, Ms. Tourangeau was appointed President of Grand Rapids Symphony in April 2005. From April 2004 to April 2005, Ms. Tourangeau served as Interim President for the Grand Rapids Symphony, leading the orchestra through its 75th Anniversary season and debut at Carnegie Hall in New York City. Ms. Tourangeau joined the Grand Rapids Symphony as Education Director in January 1997. In September 1998, she was promoted to Director of Operations. Prior to joining the Symphony, Ms. Tourangeau was the education coordinator at the Akron Symphony Orchestra, where she held that position for two-and-a-half years. She received her Bachelor of Music degree in 1994 from the Oberlin College Conservatory of Music with a major in piano performance and a minor in musicology. In April 2007, she received a Master Degree in Public Administration with an emphasis on non-profit leadership at Grand Valley State University and was inducted into Pi Alpha Alpha, the honors society for public administrators. She moved to Utah in April 2008 with her husband Michael, their 3 year-old daughter Olivia, and infant son Zachary. She is a member of the Salt Lake Rotary Club and the Alta Club, and she and her family have been enjoying all of the wonderful outdoor and family activities accessible on the Wasatch front.

2011 BEST PAPER AWARDS

Arts

Night Journey: Insight of the Subconscious of Martha Graham

Lauren Harris

Utah Valley University

Business

The Challenge: Experiential Education in Theory and Practice

Peter B. Robinson, Laurent Josien, Rachel McGovern

Utah Valley University

An Analytical View of China as a Software Outsourcing Outlet

Taowen Le

Weber State University

Wayne Huang

Ohio University

Jim Zhang

University of Wisconsin, Milwaukee

Education

Practical Theory: Making Research Writing Useful for Students

Angie Carter

Utah Valley University

Engineering

Attitude Control Using Aerodynamic Vectoring on an Aerospike Nozzle

Nate Erni, Crystal Frazier

Utah State University

Letters—Humanities, Philosophy, Foreign Language

Oedipus Family in José Luis Borau's *Furtivos* and Icíar Bollain's *Flores de otro mundo*

Matias Abeijon

Southern Utah University

Letters—Language and Literature

The Function and Fetishism of Victorian Clothing in the Neo-Victorian Novel

Amanda Scott

Dixie State College

The Museum, the Academy, and the Politics of Display: Fetishization and Aestheticization of American Indian Texts

Kimberli Lawson

Utah Valley University

Physical Science

Instantaneous Frequency Tracking of Partial of a String Vibrating at Large Amplitudes using a High-speed Camera

David C. Ripplinger, Brian E. Anderson

Brigham Young University

Social Science

Changes in Racial Concentration Ratios in Mountain States Counties, 1990–2000

L. Dwight Israelsen

Utah State University

Night Journey: Insight into the Subconscious of Martha Graham

Lauren Harris Brown

Utah Valley University

Abstract

The art of choreography is an extensive creative process informed by the artist's subconscious experiences and motivations. As one of the most famous dancers and choreographers of the modern dance world, Martha Graham had the ability to portray the subconscious through movement. The purpose of this research is to provide a deeper understanding of the creative process and the motivations that inform Martha Graham's choreographic work, Night Journey, through critical analysis of the piece and review of written source material, employing a psychoanalytical perspective.

The creative process is one of mystery. Ideas and experiences drawn from within the innermost workings of the choreographer's subconscious are molded and shaped into works of art. The epiphanies of the creative process and the motivations that enlighten them are ambiguous and unknown. It is from the deepest reaches of one's soul that intricate and exquisite—although frequently misunderstood—

expressions of humanity and culture are woven and create windows of insight into a choreographer's internal make-up. The artist turns from the realities of the world and "allows full play to his erotic and ambitious wishes" of the subconscious (Freud, 1). It is through this "play" that the choreographer faces the unknown of the human psyche to discover and express the deepest levels of human emotion. A critical analysis of Martha Graham's *Night Journey* from a psychoanalytical perspective provides delicate sapience into the intense inner subconscious motivations that inform her choreographic work.

The purpose of this research is to provide a critical analysis of Martha Graham's *Night Journey*, employing a psychoanalytical perspective to more fully understand the motivations that inform this choreographic work.

A review of written source material and a critical analysis of *Night Journey* will yield the answers to the following questions:

1. How does a psychoanalytical perspective shed light on the creative process?
2. How does a psychoanalytic evaluation of Martha Graham's life inform a discussion of her choreography?
3. How does a critical analysis of *Night Journey* from a psychoanalytical perspective provide insight into Graham's choreographic choices?

To fully understand the psychoanalytic process, it is necessary to first understand the basic concept of conscious versus subconscious. An understanding of this distinction provides greater insight into the sources of the artist's inspirations that inform the creative process as well as the choreography. As defined by Sigmund Freud, the conscious is "the conception which is present to our consciousness and of which we are aware" while the subconscious or unconscious, according to psychoanalysis, is "a conception—or any other mental element—which is now *present* to my consciousness [that] may become *absent* the next moment, and may become *present again*, after an interval, unchanged, and, as we say, from memory, not as a result of a fresh perception by our senses" (Freud, 33-4). This is to say that the subconscious is the conception of "which we are not aware" but are nevertheless able to consciously obtain and convey. Although there has been dispute over the relevancy of Freud's theories, these debates typically concern the clinical practices of psychoanalysis, which is not wholly the same as its critical application as a frame of analysis and is, therefore, still prevalent. These theories allow the viewer to become aware of the source and meaning of that which the artist is trying to convey.

It is this subconscious that the artist turns toward during the creative process of choreography. Inspiration for choreography is created as the artist takes a step back and allows "raw material emerging from the unconscious" to surface (Nachmanovitch, 9). As characteristic of modern dance, the artist, to convey the deepest levels of human emotion, must have the courage to be the one who turns away from reality and the societal constructs of what is culturally acceptable, and who then, according to Freud, "in phantasy-life allows full play to his erotic and ambitious wishes" (Freud, 1). This ability of the choreographer to consciously delve into the inner workings of the subconscious is considered by Freud to be a very "special gift" in which one "moulds his phantasies into a new kind of reality" (Freud, 1). Although there is some debate over Freud's emphasis of the importance of the erotic as an explanation of human behavior, it is through this emergence and molding of the human subconscious that the inspiration for the creative process is informed. This special ability of creating a new kind of reality is unique to the artist and is where the secrets of the subconscious of humanity and the artist himself are made known. But, the creative process and exploration of the subconscious is not something that comes to the artist and choreographer with ease. According to Freud, "it is by no means impossible for the product of unconscious [or subconscious] activity to pierce into consciousness, but a certain amount of exertion is needed for this task" (Freud, 37). He reveals that when this task is attempted by oneself, one "become[s] aware of a distinct feeling of *repulsion* which must be overcome" (Freud, 37). It is through this repulsion or resistance that one learns "that the unconscious idea is excluded from consciousness by living forces which oppose themselves to its reception" (Freud, 37). In the theories of psychoanalysis this suggests that "the repulsion from unconscious ideas is only provoked by the tendencies embodied in their contents" (Freud, 38). During the creative process, the choreographer must face his or her own subconscious resistance to allow the intricate and exquisite expressions of the inner workings of human emotion to pierce the conscious and to make known to him- or herself and the audiences the tendencies embodied within the subconscious. It is through using the psychoanalytical "special gift" during the creative process that the choreographer is able to consciously tap into the ideas and experiences drawn from within the innermost workings of their subconscious.

A psychoanalytical perspective identifies the inspiration and creativity of the choreographer's work to be rooted within the instinctual and truest material of the subconscious and, therefore, sheds light on the creative process of choreographers in the modern era such as Mar-

tha Graham. This perspective allows the viewer to become aware of the subconscious motivations that inform Graham's choreographic work.

Martha Graham was one of the great dancers and choreographers of the modern dance world. Her dance works were filled with genius that found connections between movement and human emotion. One of her principal dancers, Bertram Ross, said of her, "here was somebody who could manifest, make visible, all those feelings that you have inside you that you can't put words to" (Freedman, 12). Using Freud's psychoanalytical discussions, it is apparent that Graham had the "special gift" of consciously molding reality from the seemingly locked chambers of the subconscious. She was able to step away and allow the raw material of the unconscious that Nachmanovitch spoke of to surface and inform her choreographic process. Graham said of her process that the beginning of a new dance was "a time of great misery" as she wrote down thoughts and impressions that came to her (Freedman, 113). These thoughts and impressions that caused such misery seem to be the psychoanalytical repulsions created as the artist works to allow the tendencies and emotions embedded within the subconscious to pierce their conscious. Perhaps that curiosity and fascination of the "interior landscape" was introduced to her through her psychiatrist father, George Greenfield Graham (Freedman, 59).

Martha Graham was born in 1894 in Allegheny, Pennsylvania, the heart of the coal country. She would "remember her hometown as a bleak, gray place, 'spun entirely out of evening and dark thread'" (Freedman, 20). Graham recalls in her autobiography, *Blood Memory*, that her father's work was "a great wonder" to her as a child, and it was through him that she was taught her first lessons of truth when he told her that she "'must look for the truth' ... good, bad, or unsettling" (Graham, 19). Through these truths, she sought to bare fundamental human moods and feelings through movement. Another lesson Graham learned from her father was the idea that "movement never lies" (Freedman, 15). This lesson became the essential foundation for her view of dance and choreography, for throughout her life it was Graham's belief that "every dance is ... a graph of the heart" (Sorrell, 55). To Graham, her choreographic work could neither hide nor lie about the things locked inside, but instead were manifestations of true human character. She internalized these basic, but ground-breaking lessons into the very core of who she was. The lessons seem to have been ingrained into her subconscious, and as she created choreographic work, "she had to seek a complete identification" with the character she sought to embody; "...she had to delve into 'the depths of man's inner nature, the unconscious'" (Sorrell, 54). This complete identification is achieved by Graham through what she calls "blood memory" (Graham, 9). The memory

from “thousands of years of ... blood” is collectively carried by all and is explanatory of “instinctive gestures and thoughts” both conscious and subconscious (Graham, 10). It is through this memory inherently living within the subconscious that the “depths of man’s inner nature” is obtainable by the artist, choreographer, and Graham herself. Through digging within her own subconscious and living within her blood memory, the nature of the character she sought to embody was within reach and the complete identification she so greatly strived to achieve was able to be accomplished. A psychoanalytical discussion of the “blood memory” seems to articulate this collective subconscious as representative of the shared and instinctual tendencies that provoke repulsion. It is these instinctual tendencies that are rooted within the blood that has been shared over thousands of years and is now collective. These repulsive tendencies, such as sex and aggression, are where Graham is able to connect and identify with the character for the tendencies and instincts are shared by the whole of humanity.

As Graham started her own dance company and began to cast dancers for her roles, she looked for “a blood memory” within the dancers “in the sense that the dancer remembers and can call upon more of his or her life than has yet been lived. There has to be a hunger and a need in the dancer ... [there] has to be courage, a willingness to explore unknown feelings and daring to feel them and let them become part of your being. It’s scary. Terrifying. But you do it because you have no choice” (Horosko, 156-7). Martha knew the courage and willingness to explore unknown feelings must be present within every dancer attempting to face the resistance and repulsion put up by the subconscious. These unknown feelings are what Graham strived to share when creating her choreographic work no matter how terrifying they seemed, and she needed her dancers to be right behind her. She searched for the truth within herself and the subconscious motivations behind those truths whether they were good, bad, or unsettling, and it was these truths she and her dancers set out to convey to the audience. One of her dancers, Gertrude Shurr, said of Graham, “she had her tantrums because she couldn’t draw out of herself all of the devils she kept inside her. When she couldn’t rid herself, cleanse herself, it was just frightful” (Freedman, 109). Graham knew that the devils inside of her, the ones Shurr spoke of, were those subconscious motivations not only within herself, but also within all humankind—the collective “blood memory” (Graham, 9).

Graham felt the need to search for, recognize, accept, and share the devils she had within her with the world through her choreography, and as she did so her work began to fall into identifiable categories, one being the Greek cycle. The famous tragedies and dramas were under-

stood by the Greeks as a way to uncover the most basic and truest truths of what it means to be alive. They believed there to be a collective unconscious that every human who has and ever will live understands, and they attempted to express these eternal and shared ideas through their tragedies. Graham believed and understood this collective unconscious to be her “blood memory” and seemingly strove to uncover those most basic truths through investigating her subconscious and letting the raw material from within inform her choreography. *Night Journey* is one of these choreographic works based within a famous Greek tragedy in which Graham endeavored to express her own subconscious emotions that could and would be collectively understood.

Martha Graham’s *Night Journey* is based on the Greek myth of Oedipus Rex and Jocasta. Jocasta is queen and Oedipus is her husband. They are told by the blind seer Tiresias that they, although husband and wife, were also mother and son. When they become aware of this, Oedipus, in disgust, gouges his eyes out, and Jocasta, in shame, commits suicide. *Night Journey* opens at the end of Jocasta’s life, where she is forced by Tiresias to relive the most significant moments of her life before she is permitted the release of death.

As seen in the video *Martha Graham: In Performance*, produced by Nathan Kroll, the dance opens with Jocasta, danced by Graham, holding a silken rope high above her head with stiff outstretched fingers rotating side to side with her head back in an almost trancelike state. She is awakened at the sound of Tiresias’s wooden staff against the floor as he slowly makes his way toward her and yanks the rope from her fingers. The chorus then runs on stage to begin the reliving of Jocasta’s life. They are the Furies, the conscience of the community. They are “the terrors we all have. They are memories of things we dread to remember, things we wish to forget—the terrors. They must be recognized and lived through until they leave your mind” (Graham, 213). These horrific memories, the things we wish to forget, seem to be representative of the devils Shurr described within Graham. From a psychoanalytical perspective, these different characters seem to suggest a representation of Freud’s id, ego, and superego. The Furies, as well as Tiresias, seem to represent the psychoanalytic superego, or conscience, of Graham and Jocasta, as well as the conscience of the community. Jocasta’s want and love for Oedipus appears to be the psychoanalytic id, or subconscious, instinctual desires. Jocasta, or Graham, then becomes the psychoanalytic ego stuck between the id and the superego. They are the middle ground of the two and are the ones who must wrestle with the opposite desires of the id and the superego. Graham ven-

tures, through her use of these specific characters and their attributes, into the subconscious to convey her deepest, truest emotions.

Later, Jocasta has a solo where she moves from one side of the stage to the other kicking her leg with sharp, angular movements. The movement seems similar to a pacing motion. She stops and, in quick, jerking movements, covers the private parts of her body as she realizes the terrible sin of incest she has committed. Oedipus then enters the stage as Jocasta relives, again, his grand entrance into her kingdom. He carries her over to a stool where she watches him dance for her in a highly erotic solo as he presents his muscular legs in front of her face, almost teasing her, and wraps her in his cloak with him. Graham believed there to be “a certain beauty about sex that can only be expressed through eroticism” (Graham, 211). The eroticism displayed through the myth of Oedipus and Jocasta as well as through Graham’s choreography in Oedipus’s solo demonstrates the beauty and glory of sex and the body. Because of her view and belief about sex, Graham’s choice of this particular Greek myth in 1947 is interesting. In the theories of Freud, there is a conflict known as the Oedipus Complex, which is based upon this very myth. The theory is outlined to be that of a son who has sexual desires for his mother and sees his father to be a rival for her affections. Graham’s choice of the character, Jocasta, with her young husband, Oedipus, on whom Freud’s theory is based upon, may perhaps be, to Graham, representative of her own relationship at the time. Graham, during 1947, was also involved in a love affair with a younger man, one of her principal male dancers, Erick Hawkins. Graham’s view of the body, the beauty of sex, and the love she felt for Hawkins created a connection between herself and the character she sought to embody. The sexual relationship between Jocasta and Oedipus was shunned by the Furies, but was once sanctioned by the community when under different circumstances. Graham’s relationship with Hawkins was also something to be shunned by the community during the time period in which they were together. This inappropriate sexual relationship is similar to Freud’s Oedipus Complex. Although Graham and Hawkins were not mother and son, it was seen as a disgrace to be in love and in a relationship with a younger man in the 1940s. This relationship that could have been sanctioned by the community if under different circumstances of age allowed Graham to identify with Jocasta on a deeper, more personal level and so permitted Graham to “seek a complete identification” with the character.

There seems to be another connection when Graham states in *Blood Memory* that, “[Jocasta] has in her hands the branches that Oedipus had given her when he carried her from the bed onto her small stool where she became the queen. She drops into a wide split fall and puts

one flower out tentatively toward him, sits back and crosses her knees, opening and closing, opening and closing...It is not only a movement, but rather a gesture of invitation for him to come between her legs” (Graham, 215). Jocasta is at first hesitant, but finally gives in to the invitation to become one with Oedipus. This is similar to Graham’s personal life. Graham was once asked by the writer Joseph Campbell “if she could ever commit herself fully to love,” to which Graham replied, “If I were to take that step, I would lose my art” (Freedman, 91). Graham obviously was hesitant toward having feelings of love; she felt as though they would ruin her, but just as with Jocasta, she finally allowed and gave the invitation for Erick and herself to become one. As Jocasta, the ego, struggles between the desires of her subconscious id and the realities of her conscience superego or the conscience of the community, she finally decides to give in to the wishes of the id and invites Oedipus in. This is the same for Graham. In her personal life, she was struck with the decision between following her subconscious and instinctual desires of love for Erik and listening to the conscience of the world in which she lived. At the time, a relationship with a younger man was intensely scandalous and frowned upon. The realization of her subconscious desires may suggest Graham’s extreme loathing as she gave in to the wishes of the id and committed herself to love, thus forming a conflict between being an artist versus being a woman, for Graham had a “hard time being a woman” (Graham, 199).

After the invitation given by Jocasta and accepted by Oedipus, the two make their way to the bed and are intertwined in a sexual meeting. The Furies come again, walking slowly this time, in front of the bed with their eyes covered, disgusted by the act of incest that has taken place. They again are the conscience of the community, the terrors and memories wished to be forgotten. Graham seemingly used these Furies to represent the revulsion of her relationship. Although desperately in love with Hawkins, she must have been angry, distraught, or disgusted with herself for allowing love with a younger man to enter into her life. Perhaps she felt as though her terrors of love were coming true and her relationship with Hawkins was something that subconsciously she wished to forget.

The Furies’ distorted contracting bodies with tight, flexed muscles are a reflection of the disbelief, disgust, sorrow, and anger of the community as Jocasta and Oedipus are intertwined with the rope Jocasta held in the beginning. The rope seems to tie them together in a fate and destiny unwanted by both parties. They are separated by Tiresias, and as Oedipus realizes what he has done, he takes the rope from around his legs, holds it high above his head, and throws it to the ground in utter disgust. He runs over to the bed where Jocasta lies and slowly climbs

upon it as the Furies come in around him. He slowly raises his arms high above his head, fingers outstretched as though crying to the heavens above, and dives over the Furies' outstretched arms to lie on the floor, hiding his face as if to hide from the sin he has committed and from the eyes of the community's conscience. Oedipus makes his way again to the bed, quickly grabs the jewel from Jocasta's dress, and blinds himself before slowly staggering offstage. Jocasta then rises from the bed and very slowly walks to where the rope lies on the floor as she realizes her only escape is through death. She unfastens the robe she is wearing and lets it slide off her arms gradually to the floor. Everything from her life before falls with the robe, and she is left in a plain dress as if standing naked and vulnerable to the judgments of the sin she has committed. Jocasta then takes the rope in her hands looking at it with affection as she realizes the beautiful ugliness of her love for Oedipus and their inappropriate sexual relationship. This seems to be a reflection, again, of Graham's personal life with Hawkins. She feels a deep love and affection for Erick but is aware of the tragedy and anguish that is bound to come from their love relationship, the anguish over the censorship of the community for the inappropriate love and sexual connection she and Jocasta so profoundly want. Jocasta then turns her back toward the audience and pulls the rope around her neck. She falls to the ground with one arm still raised holding the rope. Her body tightens before she finally goes limp and there she lies, sprawled across the floor, in the "peace and forgetfulness of death" (Graham, 213).

Martha Graham had the courage to delve into the depths of her own subconscious. She had what Freud believed to be a "special gift" of consciously drawing out ideas and experiences from within the innermost workings of her subconscious to mold and shape them into works of art (Freud, 1). Graham once stated, in regard to *Night Journey*, that within the dance "there are small intimacies that I have never revealed in words. All of these things mean a tremendous amount to me" (Graham, 214). These small intimacies are representative of the issues created by her relationship with Hawkins, which she could not consciously process and so repressed, but locking away the emotions and issues she was facing did not make them go away. These issues became a part of the subconscious/unconscious that is the playground for Graham's creative output and from where she explored, through her choreography in *Night Journey*, the tension she held within. These small intimacies and locked emotions from the subconscious are shown through her choreography and allow the audience to peer into the collective conscious and "blood memory" they too hold. Graham's motivations for her choreographic choices, as seen through a

psychoanalytical perspective, were rooted within her own subconscious emotions and the conflicts of her personal life of love versus conscience and art. Through digging into the tendencies embodied within the subconscious and letting the raw material there inform her works, Graham had the tremendous ability to reach the deepest reaches of her soul as well as the souls of humanity to uncover the truest truths of what it means to live. It is through her work that one can have an insight not only into Graham and her life, but also into oneself.

Bibliography

Freedman, Russell. *Martha Graham: A Dancer's Life*. New York, NY: Clarion Books, 1998. Print.

Freud, Sigmund. *General Psychological Theory: Papers on Metapsychology/Theories on Paranoia, Masochism, Repression, the Unconscious, the Libido, and Other Aspect of the Human Psyche*. New York, NY: Touchstone, 2008. Print.

Graham, Martha. *Blood Memory/Martha Graham*. New York: Doubleday, 1991. Print.

Horosko, Marian. *Martha Graham: The Evolution of Her Dance Theory and Training*. Gainesville, FL: University Press of Florida, 1991. Print.

"*Martha Graham: In Performance*". Produced by Nathan Kroll. Conceived and choreographed by Martha Graham. Kultur International Films, 1989. Film.

Nachmanovitch, Stephen. *Free Play: Improvisation in Life and Art*. New York, NY: Penguin Putnam, 1990. Print.

Sorell, Walter. "Martha and Myth." *Dance Magazine* 1 July 1991: 53-56. Print.

Batrachochytrium dendrobatidis **Fungus Update for *Hyla arenicolor* in Zion National Park**

Britnee Eng and Curt Walker
Dixie State College

Abstract

Batrachochytrium dendrobatidis is a form of chytrid fungus that results in a condition known as chytridiomycosis. It is a contributing factor in decreasing global amphibian populations. To date, chytridiomycosis has been found on all continents and in various areas within the United States. Data from the summer of 2009 indicated that chytridiomycosis was not found in the canyon tree frog, *Hyla arenicolor*, within Zion National Park in southwestern Utah. To determine whether these populations were still free of the chytrid fungus one year later, samples were collected in the summer of 2010 and tested for the presence of *B. dendrobatidis*. Of the 10 canyons from which epidermal swab samples were collected, three showed positive results for the presence of *B. dendrobatidis*. These results can help us establish a timeline for the spread of *B. dendrobatidis* into the amphibian populations of Zion National Park.

Introduction

Chytridiomycosis is a disease caused by the fungus *Batrachochytrium dendrobatidis*; it has been shown to be a major cause of amphibian decline around the world (Berger et al. 2005) and has been linked to the extinction of at least 30 frog populations (Kilpatrick et al. 2009). Chytrid fungus is believed to have originated in South Africa in the 1930s (Weldon et al. 2004). It is presumed to have spread because of the increase in the trade of *Xenopus laevis* specimens in 1934 for their use in the human pregnancy assay (Weldon et al. 2004) as well as the increase in the international pet and zoo trade (Kilpatrick et al. 2009).

Chytrid fungus appears to spread via skin-to-skin contact or by contact with infected water (Forzan et al. 2008). Typical signs of infection in amphibians include excessive sloughed skin, skin discoloration, lethargy, hind legs held loosely behind the body (resulting in an abnormal sitting posture), and lack of predation avoidance behaviors (Berger et al. 2005). Death usually occurs within 18 to 70 days after initial infection (Berger et al. 2005), but varies greatly from species to species.

The exact mechanisms that make *B. dendrobatidis* lethal to most species of amphibians are not fully understood (Kilpatrick et al. 2009). Some amphibians show a strong innate immune response that makes them less vulnerable to the fungus; environmental conditions, such as overcrowding (Woodhams et al. 2007), may have a considerable effect on this immune response (Carver et al. 2010). *B. dendrobatidis* affects most amphibian species after zoospores penetrate the epithelial cells of an individual and then use the keratin as a source of nutrition (Kilpatrick et al. 2009). Damaged epithelial cells can affect respiration and electrolyte levels in infected frogs, as these processes normally are regulated through the frog's semi-permeable skin (Berger 1999; Raverty and Reynolds 2001; Rollins-Smith et al. 2002). Additionally, rates of obtaining necessary water after dehydration in infected frogs have been shown to be drastically affected by *B. dendrobatidis* and may be the main reason for fatality in most amphibian species (Carver et al. 2010). Tadpoles, with their reliance on lungs for gas exchange, are typically immune to the effects of the fungus but may be vectors or reservoirs of the disease (Bosch et al. 2001; Van Ells et al. 2003). Chytrid fungus can often be detected affecting the keratinized mouthparts of tadpoles, but this is not apparently harmful to the tadpoles.

Previously, populations of *Hyla arenicolor* in Arizona have been shown to be infected by the chytrid fungus *B. dendrobatidis*, and a small number of deaths may have been associated with this infection (Bradley et al. 2002). *H. arenicolor*, also known as the canyon tree

frog, is found across most of the southwestern United States (Bonine et al. 2008). It can be found living in pools in slot canyons or in rocky areas along a water source. During the summer, *H. arenicolor* basks on rocks during the daylight, perhaps to avoid snakes or to reduce parasite loads (Cagle 1950).

Zion National Park (ZNP) is close to the affected populations of *H. arenicolor* in Arizona; we decided to monitor canyons in ZNP for the presence of the chytrid fungus to establish a timeline for the possible spread of the fungus. Previous studies carried showed two positive results for *B. dendrobatidis* in Pine Creek canyon (K. Bonine, unpublished data), but research carried out in 2009 showed negative results for the presence of *B. dendrobatidis* in populations of *H. arenicolor* in ZNP (Jones et al. 2010).

To determine whether *B. dendrobatidis* had reached populations of *H. arenicolor* within ZNP since the summer of 2009, samples were collected from *H. arenicolor* individuals in canyons in and surrounding ZNP in 2010. The samples were sent to the San Diego Zoo Institute for Conservation Research and analyzed there for the presence of *B. dendrobatidis*.

Materials and Methods

Sample Collection

We were permitted to sample from 10 different slot canyons within ZNP, most of which are popular hiking destinations. We chose to sample from locations that would provide a range of population sizes and a range of degrees of isolation. Pine Creek is near the center of human activity in the park, drains into the Virgin River, and has a frog population of hundreds or even thousands; Fat Man's Misery is rarely visited, quite remote, and has a small population of frogs, certainly less than 50 adults. By sampling from a range of canyons, we sought to discover whether chytrid was more likely in commonly visited locations and whether population size influenced the presence of chytrid. Sampling locations are listed in Table 1.

We collected epidermal swab samples from 10 frogs within each slot canyon each time we entered the canyon; if fewer than 10 frogs were found, samples were collected from all encountered frogs, otherwise individuals were chosen at random as they were encountered. Epidermal swab samples were collected from 72 individuals from 10 different locations in and surrounding ZNP during May to July of 2010 (Figs. 1 and 2; Table 1). The capture and swabbing techniques used were described by Jones et al. (2010).

Table 1. List of canyons where epidermal swab samples were taken from as well as the frog population count and the number of samples obtained.

Slot Canyon	Hyla arenicolor Population Count	Epidermal Swabs Obtained
Leap Creek	4	3
Taylor Creek	21	11
Keyhole	18	4
Checkerboard Mesa Pools	9	9
Fat Man’s Misery	12	10
Pine Creek	30	10
Many Pools	48	10
Russell Gulch	9	6
Subway	86	9
Lower Russell Gulch	-	5

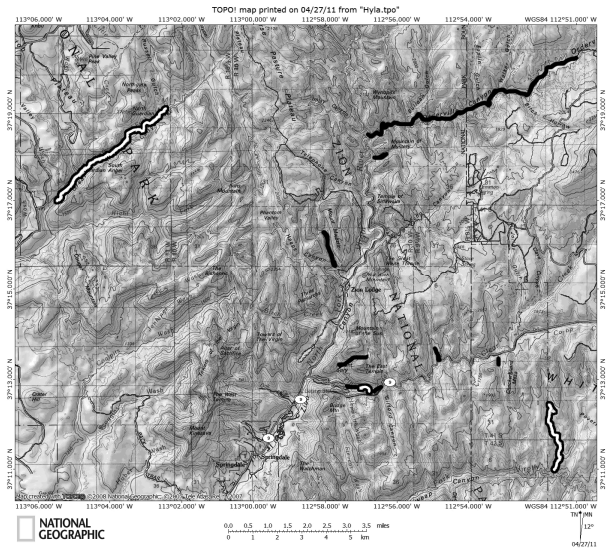


Figure 1. Map of main canyon of Zion National Park and the Subway; study canyons are indicated by thick black lines and canyons where epidermal swab samples tested positive for *Batrachochytrium dendrobatidis* are indicated by white lines. (Map produced with National Geographic TOPO!)

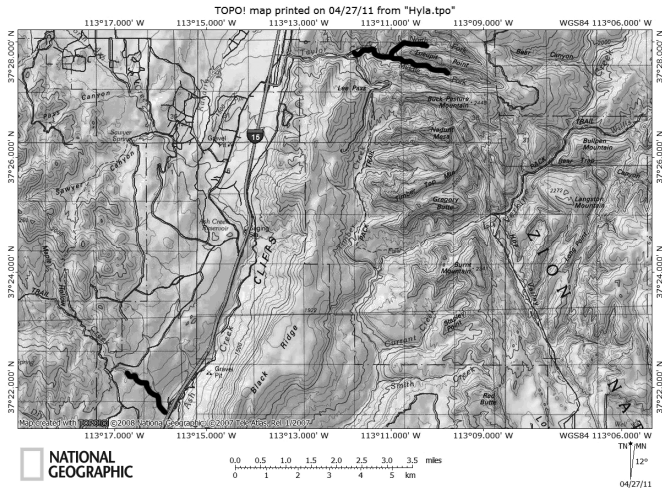


Figure 2. Map of northwest Zion National Park; canyons are indicated by thick black lines. No canyons in this area tested positive for *Batrachochytrium dendrobatidis*. (Map produced with National Geographic TOPO!)

PCR Analysis

Each sample was sent on dry ice to the Amphibian Disease Laboratory at the San Diego Zoo Institute for Conservation Research. Upon arrival at the Institute, samples were kept frozen at -20°C until analysis. The amphibian swab sample DNA was extracted using Prepman Ultra Sample Preparation Reagent (Applied Biosystems, catalog # 4318930), followed by a chytrid fungus real-time PCR assay, using Universal Taqman master mix, to test for the presence of *B. dendrobatidis* DNA. The real-time PCR was used to amplify the ITS-1/5.8S junction of the zoospore DNA. Chytrid DNA was detected, if present, by using primers and probes described by Boyle et al. (2004). Positive results were reported whenever any *B. dendrobatidis* DNA was detected in a sample well.

Results

Of the 10 locations that were sampled, three showed positive results for the presence of *B. dendrobatidis* (Table 2). Chytrid fungus was found in four of the samples from Pine Creek, with the positives all found in lower sections of the canyon. Chytrid was also detected in four samples from Fat Man's Misery Canyon, West Fork, which is techni-

cally outside the park boundaries to the southeast, and seven positives in the Left Fork of North Creek (known as the Subway), in the western section of Zion.

Table 2. Results from PCR testing for presence of <i>Batrachochytrium dendrobatidis</i> in epidermal swab samples		
Amphibian collection location	Number of animals collected	Number of animals testing positive
Checkerboard Mesa	9	0
Fat Man's Misery	10	4
Keyhole	4	0
Leap Creek	3	0
Lower Russell Gulch	5	1
Many Pool's	10	0
North Fork Taylor Creek	10	0
Russell Gulch	6	5
Subway	4	1
Taylor Creek	1	0
Pine Creek	1	4

Discussion

The distribution of the fungus within the greater ZNP area is perplexing, with no obvious pattern to explain the distribution of infected frogs. Fat Man's Misery Canyon is rather rarely visited, and yet several samples from Fat Man's are chytrid positive. Fat Man's Misery is usually visited by persons starting at Checkerboard Mesa, but requires a long, difficult hike of several miles; yet the pools at the upper part of the drainage have the fungus. Pine Creek and the Subway are common tourist destinations and have the fungus, but Taylor Creek and Many Pools canyons are also commonly visited, and they are chytrid-free according to our data thus far. The pools at Checkerboard Mesa are only a few yards from the main East–West highway through the park, and they appear to be chytrid-free. A feasible explanation for this distribution is that the fungus was introduced on separate occasions by hikers, and the viable spores have only become established in Pine Creek, the Subway, and Fat Man's Misery canyons.

Dead adult frogs are rarely encountered in our excursions, despite our active searching for them and the occurrence of the fungus in some of our study canyons. Kriger and Hero (2007) found that chytrid infection rates in Australia were best predicted by habitat type and adult frog

size; large adults in drier, warmer habitats were least likely to suffer chytrid mortality. It appears, therefore, that some affected species in dry climates (which likely depend very little on cutaneous respiration) were able to maintain reasonably high populations despite the presence of chytrid fungus. Adult canyon tree frogs most likely don't rely heavily on cutaneous respiration, as they allow their skin to dry out for a substantial portion of each day during the desert summer and can tolerate rather extensive loss of body water (Cunningham 1964; Wylie 1981). Perhaps its habit of perching in the sun in an arid climate protects *H. arenicolor* adults by inhibiting fungal growth. Snyder and Hammerson (1993) found that canyon tree frogs maintained body temperatures between 29 and 31°C, too warm for chytrid fungus (Berger et al. 2005).

During the fall, canyon tree frogs begin brumation in cracks among rocks, tolerating freezing conditions quite regularly, until the weather warms in the spring (Feder and Burggren 1992). Unless they depend heavily on cutaneous respiration at this time, which could be fatal in the presence of the fungus, this freeze-tolerant overwintering habit could help to make them relatively resistant to the effects of chytrid infection over the winter. Other researchers have found that chytrid mortality was seasonal (Retallick et al. 2004; Voordouw et al. 2010), with the lowest infection rates during hot summers (the fungus is most likely inhibited at temperatures above 28°C; Longcore et al. 1999). There may be a season during which our study species is particularly vulnerable, but so far cautious optimism seems justified.

Resistance to the fungus could explain why dead frogs are rarely encountered in ZNP. Rollins-Smith et al. (2002) and Woodhams et al. (2007) found that, while the fungus was typically out of reach of immune system cells, antimicrobial proteins commonly found in hylids were sometimes potent killers of the fungus. In *Rana pipiens*, some individuals have been found to be resistant to *B. dendrobatidis*, and they were able to clear it completely from their systems (Voordouw et al. 2010). Similarly, Retallick et al. (2004) found that chytrid resistance was beginning to appear in an Australian rainforest frog species. With that in mind, we hope that our study species, *H. arenicolor*, will be another example of a chytrid-resistant organism. Further studies will be carried out on *H. arenicolor* in ZNP to determine whether this species does show chytrid fungus resistance.

We plan to continue monitoring populations of the frogs in these study canyons and track the spread of the fungus in the area. Monitoring of populations in the study canyons located in ZNP began in the summer of 2008; unfortunately, we have been unable to design an effective protocol for mark/recapture studies. Only raw counts of ob-

served frogs on each hike have been obtained; this is shown to be less accurate than mark/recapture studies (Seber 1965, 1982, 1986). Although these raw counts are likely to vary with weather conditions, the inevitable variations in time of day and spotting success, they should allow us to maintain a rough approximation of population sizes, and make substantial population drops apparent. Raw population counts will also allow us to determine whether the canyon tree frog is vulnerable to *B. dendrobatidis* and the degree of vulnerability. Sample collection will also continue from these research areas to determine whether the fungus spreads, whether it is ever eliminated from canyons once it has been introduced, and whether the fungus proves to be detrimental to populations of *H. arenicolor* in those canyons where it is present.

References

- Berger, L., Speare R., and Hyatt, A. 1999. Chytrid fungi and amphibian declines: Overview, implications and future directions. In: *Declines and Disappearances of Australian Frogs*. A. Campbell (Ed). pp. 21–31. Environment Australia, Canberra.
- Berger, L., Hyatt, A.D, Speare, R., and Longcore, J.E. (2005). Life cycle stages of the amphibian chytrid *Batrachochytrium dendrobatidis*. *Dis Aquat Organ*, 68: 51-63.
- Bonine, K., Edwards, T., Schirmer, C., Fillinger, J., and Ratzlaff, K. (2008). Distribution and phylogeography of the Canyon Treefrog, *Hyla arenicolor*, in the Rincon Mountains, Arizona. *Sonoran Herpetologist*, 21(10):106-109.
- Bosch, J., Martínez-Solano, I., and García-París, M. (2001). Evidence of a chytrid fungus infection involved in the decline of the common midwife toad (*Alytes obstetricans*) in protected areas of central Spain. *Biol Conserv*. 97(3):331-337.
- Boyle, D. G., Boyle, D.B., Olsen, V., Morgan, J.A.T., and Hyatt, A.D. (2004). Rapid quantitative detection of chytridiomycosis (*Batrachochytrium dendrobatidis*) in amphibian samples using real-time Taqman PCR assay. *Dis Aquat Organ* 60: 141-148.
- Bradley, G.A., Rosen, P.C., Sredl, M.J., Jones, T.R., and Longcore, J.E. (2002). Chytridiomycosis in native Arizona frogs. *J Wildlife Dis*, 38(1):206-212.

- Cagle, F.R. (1950). The life history of the slider turtle, *Pseudemys scripta troostii*. *Ecol Mongr*, 20: 31-54.
- Carver, S., Bell, B.D., and B. Waldman. (2010). Does chytridiomycosis disrupt amphibian skin function? *Copeia*, 3, 487-495.
- Cunningham, J.D. (1964). Observations on the Ecology of the Canyon Treefrog, *Hyla californica* *Herpetologica* 20:55-61
- Feder, M.E. and Burggren W.W. (1992). *Environmental physiology of the amphibians*. The University of Chicago Press, Chicago.
- Forzán, M.J., Gunn, H., and Scott, P. (2008). Chytridiomycosis in an aquarium collection of frogs: diagnosis, treatment and control. *BioOne*. 39(3): 406-411.
- Jones, A.D.; Sumko, D.J., and Walker C. 2010. Emerging infectious disease Chytridiomycosis and the significance of the absence of *Batrachochytrium dendrobatidis* on *Hyla arenicolor* in Zion National Park. *J. Utah Acad. Sci Arts Letters*, 87: 49-56.
- Kilpatrick, A.M, Briggs, C.J, and Daszak, P. (2009). The ecology and impact of chytridiomycosis: an emerging disease of amphibians. *Trends Ecol Evol*, 25(2): 109-118.
- Kruger, M.K, and Hero, J.M. (2007). The chytrid fungus *Batrachochytrium dendrobatidis* is non-randomly distributed across amphibian breeding habitats. *Divers Distrib*. 13(6): 781-788.
- Longcore, J.E., Pessier, A.P., and Nichols, D.K. (1999). *Batrachochytrium dendrobatidis* gen. et sp. nob., a chytrid pathogenic to amphibians. *Mycologia*. 91(2): 219-227.
- Raverty, S and T. Reynolds. (2001). Cutaneous chytridiomycosis in dwarf aquatic frogs (*Hymenochirus boettgeri*) originating from southeast Asia and in a western toad (*Bufo boreas*) from northeastern British Columbia. *Can. Vet. J.* 42(5): 385-386.
- Retallick RWR, McCallum H, and Speare R (2004). Endemic infection of the amphibian chytrid fungus in a frog community post-decline. *PLoS Biol* 2(11): e351.
- Rollins-Smith, L.A., Doersam, J.K., Longcore, J.E., Taylor, S.K.,

Shamblin, J.C., Carey, C., and Zasloff, M.A. (2002). Antimicrobial peptide defenses against pathogens associated with global amphibian declines. *Dev. Comp. Immunol.* 26(1): 63-72.

Seber, G.A.F. 1965. A note on the multiple recapture census. *Biometrika* 52: 249-259.

Seber, G.A.F. 1982. *The Estimation of Animal Abundance and Related Parameters*. 2nd ed, MacMillan, NY, NY, USA.

Seber, G.A.F. 1986. A review of estimating animal abundance. *Biometrics* 42: 267-292.

Snyder, G. K., and G. A. Hammerson. (1993). Interrelationships between water economy and thermoregulation in the canyon tree-frog *Hyla arenicolor*. *J. Arid Environ.*, 25:321–329.

Van Ells, T., Stanton, J., Strieby, A., Daszak, P., Hyatt, A.D., and Brown, C. (2003). Use of immunohistochemistry to diagnose chytridiomycosis in dyeing poison dart frogs (*Dendrobates tinctorius*). *J. Wildl. Dis.* 39: 742-745.

Voordouw MJ, Adama D, Houston B, Govindarajulu P, and Robinson J. (2010). Prevalence of the pathogenic chytrid fungus, *Batrachochytrium dendrobatidis*, in an endangered population of Northern Leopard Frogs, *Rana pipiens*. *BioMed Cent Ecol.* 10:6

Weldon, C., Preez, C.U., Hyatt, L.H., Muller, A.D., and Speare, R. (2004). Origin of the amphibian chytrid fungus. *Emerg. Infect. Dis.*, 10:2100-2105.

Woodhams, D.C., Vredenburg, V.T., Simon, M.-A., Billheimer, D., Shakhtour, B., Shyr, Y., Briggs, C.J., Rollins-Smith, L.A., and Harris, R.N. (2007). Symbiotic bacteria contribute to innate immune defenses of the threatened mountain yellow-legged frog *Rana muscosa*. *Biol. Conserv.*, 139:390-398.

Wylie, S. R. (1981). Effects of basking on the biology of the canyon treefrog, *Hyla arenicolor*. Ph.D. Dissertation, Arizona State University, Tempe, AZ.

“Right to Work”: Has It Become the “Right to Fire”?

An Editorial Review

Troy Callahan

Utah Valley University

Abstract

This paper explores whether a misconception exists about the right-to-work laws in Utah and whether that general misunderstanding contributes to abuse by employers of their employees. There seems to be a tacit acceptance by Utahans to comply with authority, and this includes religious leaders as well as employers. This sunny disposition and earnest desire to support their local leaders combined with a misunderstanding of right to work makes it a fertile ground for potential abuse. In addition, the state of the economy contributes to an acceptance by employees to endure whatever their employer dishes out. The economy, and employees willing to do anything on the job, should be an opportunity for employers to develop a loyal and long relationship with their employees, but the perception that employees can be terminated without cause leads employers to “burn” through employees without regard to how it affects their business or their employees.

Introduction

“Right to work” is three words that seem to mean something different to many people. I recently asked a coworker what he understood these words to mean, and he replied that it meant he wasn’t entitled to any breaks and that our boss could fire us anytime he wanted. I was not satisfied with that answer. I knew that there was more to right to work than just hiring and firing. I decided to go to a source that was unbiased to find out what these three words meant. I wanted to ask someone who could give an objective opinion without interjecting personal feelings or experiences; so I asked my 14-year-old son. He thought for a few moments then replied, “I guess it means that I have the right to work, if I want to.”

My son’s interpretation of the meaning of “right to work” seemed more in line with my own thinking, and I wanted to explore the issue further. I decided to take an impromptu poll of various people I meet in my everyday life to get their opinion of what “right to work” means to them. I asked coworkers, classmates, waitresses, checkout clerks, friends, and relatives what their opinion was of right to work. Most of them had the same idea that it meant our employers had the right to fire us without cause. I was surprised that the issue of unions never came up. “At-will employment” is a contract between employers and employees that exists in numerous states, and it seems to get confused with the right-to-work laws. Some people I polled even think that at-will employment is a part of the right-to-work law.

Many employers have a misunderstanding of the right-to-work law, or they simply choose to allow their employees to believe that the law simply allows them to terminate employment anytime and for whatever reason. This has created an unfair advantage in favor of the employer. Employees, who believe that they have no rights, are unlikely to report incidents of abuse or unfair treatment because they are afraid that they will lose their job. People in certain demographics suffer even more because their ability to change jobs is limited. Early in the industrial revolution, and before the influence of labor unions, workplace safety, fair wages, and employee rights were nonexistent. The only recourse employees had to protect themselves was to simply quit (Kaufman, 1998). Today, quitting or being fired is still the prevalent recourse used to solve employment issues. For employers and employees, right to work simply means, right to fire.

Background

Right-to-work laws originally addressed union shops. These laws prohibited closed shops, meaning employers cannot refuse to hire non-

union members, and union members cannot be refused work. In other words, union member or not, all people have a right to work. States without right-to-work laws have been called “right-to-good-wages-and-safe-working-condition” states, and right-to-work states have been called “right-to-fire” and “right-to-make-low-wage” states.

In 1955, the state of Utah enacted right-to-work laws. Previously, Church of Jesus Christ of Latter-Day Saints (L.D.S.) President Heber J. Grant referred to unions or closed shops as “Satan’s clubs” that take away a man’s right to work. In 1965, L.D.S. President David O. McKay affirmed the Church’s position saying that unions infringe on man’s “free agency.” Right-to-work laws were meant to allow men to work regardless of their union affiliation, but they also permit men to join a union if they desire. Arguably, the L.D.S. Church is very closely linked to Utah’s government. In fact, L.D.S. leader David O. McKay sent a letter to the 12 Mormon congressmen who were serving in 1965, asking them to vote against efforts to repeal right-to-work policies in the U.S. When one of the congressmen suggested that it was an intrusion by the church to involve itself in politics, he was assured that President McKay’s letter was just an opinion. Nevertheless, after the congressman voted in favor of repealing the law, he was defeated in the following election (Davies, 2011). In the early development of Utah, union shops were common. In fact, the idea of union shops was brought to Utah by early members of the L.D.S. Church who belonged to closed shops before coming to America. The involvement of the L.D.S. Church in passively influencing its members in political issues may be one of the reasons why “right to work” is a concept that is misunderstood, but vigorously defended, by Utahans. When the U.S. military was searching for a site to develop its chemical and nuclear weapons, Utah was chosen because the military believed that Utahans had a natural deference for authority and that they would not sue over accidents or disasters (Seegmiller, 1988). This quality seems to be one of the reasons why employers are able to exert authority and control over their employees by simply citing “right to work” as the law that trumps all others when it comes to employment issues. When combined with at-will employment agreements, “right to work” takes on a meaning that has been distorted by employers in such a way that it encourages or facilitates some employers to abuse and intimidate their employees.

Threats of termination and demotions are used to motivate employees rather than incentives and rewards. This negative reinforcement is creating an atmosphere of mistrust and fear that can only hurt employers in the long run. How can an employer engender loyalty from employees if they continually threaten to fire their employees? Employers who regard their employees as assets and treat

them as such have more success and are more profitable in the long run.

The cost of high turnover in business should be recognized by employers. To constantly hire and retrain employees puts a financial drain on companies. Companies with educated management are less likely to do this than small business owners with little or no education.

Gender and Right to Work

Several weeks ago, I was waiting in line to purchase a bottle of water from a convenience store. I was waiting behind a man who was attempting to hand the clerk a \$10 bill to buy gas. Before the clerk would accept the ten dollars, I overheard her pestering the poor man to buy stamps, a beverage, potato chips, a big pink cookie, etc. He politely declined each solicitation and was finally able to hand her his ten dollars. When it was my turn, the clerk took a deep breath, but before she could begin her sales pitch, I said, "Young lady, I am not interested in buying stamps, a soda-pop, potato chips, or a big pink cookie; I only want to buy this bottle of water!" The young woman's shoulders slumped, and she said, "If I don't ask you to buy those things my boss said he would fire my ass. This is a right-to-work state." She looked a little down, and she continued, "I'm a single mother, and I need this job."

Women are especially susceptible to abuse in the workforce, and an employer's misrepresentation of right-to-work laws makes it even more difficult for women and single mothers to feel confident in the workplace. The hospitality industry utilizes a large number of women in its business. Hotel maids, waitresses, and cleaning women make up the majority of the workforce in this industry. Even in the face of legislation and existing unions, managers still feel free to exert coercive managerial control over these workers. They feel unconstrained in hiring and firing at their discretion. One of the reasons may be that immigrants make up a large proportion of this workforce, and their understanding of employment rights is limited (Head and Lucas, 2004).

The woman from the convenience store did not even mention union membership. Her total understanding of right to work was that she lives in a "right-to-fire" state. How many women endure this type of intimidation on a daily basis in Utah? I made it a point to interview 10 women from different backgrounds with different education levels, and 100% of them thought that "right to work" means "right to fire." In fact, the men that I have questioned said the same thing. "Right to work" means "right to fire."

Men experience abuse in the workforce on a different level than women do. For men, employment is closely associated with their self-esteem and worth in the community. Having a job is a symbol of masculinity, attractiveness, and self-worth. For unskilled laborers and minorities, employment is the goal. These workers are less concerned with job fulfillment and happiness, and they concentrate more on keeping their job. Many employers recognize that their employees have no place to go and use this to their advantage.

At-will employment has its roots in British common law. The intention of the law was to allow free agents, employees, and employers to terminate employment anytime and without cause. This was meant to allow employees to take better-paying jobs or to allow employers to move employees along who might be past their prime or who couldn't contribute effectively. Today, employers seem to be the only ones who benefit from this law. Professional athletes who become free agents can do well shopping their talents around, but for the most part, employers are the real benefactors of at-will employment (Mello, 2002). Uneducated and unskilled men who have moved past their physical prime, or have just gotten older, can become victims of the right-to-fire mentality of some employers. Like their female counterparts, these men take jobs that are low paying and, in some cases, demeaning. This "last stop" in the final chapter of their working years before becoming Social Security recipients can be hard and cruel. Employers who employ these workers know that they can take advantage of them and not worry about complaints or issues arising from long hours or low wages. Employers may not be breaking any laws, but the fear of termination keeps these workers in line. Older employees without skills or education are less likely to rock the boat, and without labor unions advocating for them, they simply accept the working conditions handed to them. These older employees have the right to work regardless of their age, and their experience can be a valuable asset to an employer.

Employers

There are large and small businesses in Utah, and they all have different levels of education and sophistication. I worked for a company in Utah that was relatively small. The company had 12 employees, including me. The organization of the company consists of a husband and wife who are the owners, a general manager, two secretaries, five technicians, an estimator, and a superintendent. The general manager is a young man with no formal education. He is 26 years old and is eager to tell everyone that he makes \$80,000 per year. He also sexually harasses the secretaries on a daily basis. During one produc-

tion meeting, he stopped in mid-sentence and said, pointing to one of the secretaries, "Doesn't she have nice boobs?" We all looked at each other in amazement, but nobody spoke up to defend the secretary. After the meeting, I asked the secretary what she was going to do. She replied, "Nothing." I went to the owner of the company and told him what happened. He just smiled and said that he would speak to the general manager. We never saw any improvement in the treatment of the secretary by the general manager.

There are many employers who feel they have the freedom to treat employees any way they wish because of their faulty understanding of right-to-work laws. Most know that they cannot sexually harass employees, but they also know that unless there is a complaint, there can be no consequences to their actions. Some employers take advantage of a poor economy and their employees' fear of losing a job to keep them in line. They don't lead their company; they drive it with fear and intimidation.

Cultural Biases

In Utah, the acceptance of authority figures compounds the problem because many of these business owners are also church leaders who serve in a position of authority in the ecclesiastical matters of some of their employees. It is not uncommon for some L.D.S. Church leaders to run large businesses that employ hundreds of employees. This relationship between employer and employee now has an added component of employer/church leader and employee. The right-to-work law as it pertains to employee rights is never considered. How can an employee who also belongs to a church, and is admonished to accept their church leader as the absolute authority over their spiritual life, not have that attitude if their church leader also happens to be their boss? I am not suggesting that church leaders intimidate flock members who also happen to work for them, but the situation has the potential for abuse.

Since "right to work" became a law in Utah in 1955, a generation has been raised either to understand the law correctly or to be taught their parents' misunderstanding and biases of the law. I had the opportunity to participate in a classroom exercise facilitated by my Human Resources professor. As a class, we were divided into two groups. One group would act as company management, and the other group would act as union management. We were given a scenario to negotiate, and a point system would be used to determine who was "winning" the process. After several attempts, the points were added up, and the company managers had bested the union managers. During the exercise, I found

it interesting that the students who played the role of the company managers were totally inflexible. Any efforts to meet in the middle by the union managers were shot down by the company managers. Most of these students were young, in their late teens or early twenties. I know that most of them had no real-world experience with union management, and I was amazed at how antagonistic they were towards the union. I can only guess that their understanding of unions was taught to them by their parents or other leaders.

These young people will manage businesses in the future here in Utah, and many of them will simply perpetuate the misunderstanding of right-to-work laws. In the classroom exercise, I heard many of these young students simply say, "Let's just fire them!" A misunderstanding of a law or moral issue can become ingrained in the culture if it is taught to generation after generation. In my opinion, a hatred for union labor, and the misunderstanding that the "right to work" does not mean the "right to fire," is becoming a cultural bias in Utah. Most Utahans are not small business owners, and the majority of them currently benefit from the political efforts of unions in the past to push for social reform in areas such as safety and higher wages (Palley and LaJeunesse, 2007). When the minority is able to convince the majority to accept any issue as a doctrine or law, it becomes difficult if not impossible to make changes for the better. The uninformed and antagonistic attitude towards unions that is being taught to young people is contributing to the notion that it is morally acceptable to simply fire employees at any time and for any reason. It is also creating a timid workforce that finds it daunting to speak up and advocate for themselves or others. This also keeps businesses from gaining productive and loyal employees who want to grow with a company because they want to be there as opposed to being there because they have no other choice.

Lately, I continue to hear from people that Utah has the best economy in the country. The wages are low here, nepotism is rampant, cronyism is the business norm, and employees are constantly told that they can be fired any time. This culture has been effectively turned against unionism in favor of nepotism and cronyism. These people with the sunny dispositions have been taught to accept their low-paying jobs and abusive work environments with a smile; otherwise, they can be fired any time! Utah advertises itself as being "friendly" to small businesses. This friendliness can be at the expense of the working people in Utah when the featured benefit to doing business in Utah is the right to fire. In my opinion, the Chamber of Commerce description of Utah should be something loftier than that.

Conclusion

The right-to-work law in Utah has been effectively misrepresented to mean the “right to fire.” Uneducated and less-sophisticated small business owners have allowed themselves and their employees to believe that the main component of the law is the right of the employer to terminate his employees without cause. The employee does not know that he does have a right to organize or join a union. The employee does have the right to pay union dues if he/she chooses, and they cannot be fired for “any reason.” Young people in Utah are being taught from an early age by their parents or other leaders that union organizations are bad, even though they currently benefit from positive efforts by unions like minimum wage and workplace safety programs. At-will employment gives an opportunity for employees as well as employers to seek out the best possible situation to improve their position without worrying that their current employment will not allow them to move on. The law has been twisted to mean that employers can fire their employees without cause or justification, and this is creating an environment that is conducive to abuse by employers. Because this is coupled with a bad economy, employees are afraid to speak up and advocate for themselves for fear of termination. This, in turn, is creating a timid workforce that will not only fail to speak out against abuse but also has a fear to speak up for any reason including suggesting ways to improve a company. Unions make it difficult for a company to get away with unsafe working conditions and low wages. They provide more to employees than just a job. Employers benefit from right-to-work laws because they can keep wages down, and that keep profits up. Unions can hurt small businesses by organizing and pushing wages up, but a positive tradeoff is confident employees that take ownership in the company, which creates loyalty on both sides. Employees that constantly feel threatened will eventually withdraw from the company psychologically and then physically, which will hurt the company. It makes sense to turn to the understanding my son has of the right-to-work law. As people, we have a right to work, and companies have the right to have happy, loyal employees working for them of their own free will, without fear and intimidation. Smart companies will engender loyalty and creativity from their employees by making them feel valued and wanted. This working environment can propel businesses to productivity and higher profits because people will do more for their employers when they feel secure, not threatened. This means that employers need to understand right-to-work laws and help their employees understand them too. Employees have an obligation to their employers as well. They need to provide a good product that can be valued and appreci-

ated by their employer. It is not enough to just “show up.” As employees, we need to be engaged and enthusiastic.

We all need to remember that the rights in the right-to-work law apply equally to employees and employers. A proper and honest interpretation of the right-to-work law is an obligation that employers have to their employees and communities. Good companies develop and maintain good reputations in their local communities, which will give them sustainability in the long run.

Bibliography

Davies, J. K. (2011) “Right-to-work movement.” Utah History Encyclopedia. Available at <http://www.media.utah.edu/UHE/r/RIGHTTOWORK.html>. Accessed November 8, 2011.

Head, J., Lucas, R. (2004). “Does individual employment legislation constrain the ability of hospitality employers to “hire and fire?” *International Journal of Hospitality Management* 23 (3), pp. 239-254.

Kaufman, B. E. (1998 March). “Regulation of the employment relationship: The “old” institutional perspective.” *Journal of Economic Behavior and Organization*, 34 (3), pp. 349-385.

Mello, J. (2002). “Employment-at-will vs. wrongful discharge.” *Business Horizons*, 45 (6), pp. 3–6.

Palley, T. I., LaJeunesse, R. M. (2007). “Social attitudes, labor laws, and union organizations: Towards a new economics of union density.” *Journal of Economic Behavior and Organization* 62 (2), pp. 237-254.

Seegmiller, J. B. (1988). “A History of Iron County: Community Above Self.” Utah State Historical Society.

Teaching Theory Versus Practical Use: The Case of the Modern Portfolio Theory

Leo H. Chan

*Utah Valley University and Tan Tao University,
Vietnam*

Abstract

In this teaching note, I demonstrate how literal application of the Modern Portfolio Theory (MPT) could lead to inconsistent performance by using data from the U.S. stock markets. The demonstration shows that it is impossible to construct a forecast of an efficient frontier by static analysis of the MPT. While the minimum variance portfolio does produce better risk adjusted return for the investor during the 1996 to 2005 period, the result from 1987 to 1997 period show otherwise. Therefore, using any of the tools derived from MPT without taking into account the time-varying nature would produce inferior results for individual investors.

I: Introduction

There is an idiom that says that old habits die hard. It is even more so in academia, where some well-established theories continue to garner more coverage in textbooks than the counterevidences that follow. The purpose of this teaching note is to point out how a well-established theory in finance that works well with historical data might not be working so well when it matters the most: forecasting performance. I chose the Markowitz Portfolio Theory (Markowitz, 1952), or Modern Portfolio Theory (MPT), because it has been the cornerstone of modern finance for many decades and formed much of the portfolio management practices common in developed markets around the world. Many studies have since been published to demonstrate the shortcomings of the MPT, yet most textbooks still devote significantly more coverage to MPT than any of the studies that demonstrated how impractical the MPT is. It is as if all textbook authors decided to treat the doubts about MPT as an afterthought at the same time. I believe the practice of continuing the teaching of an established theory without balancing with the counterevidence is a hindrance to better understanding of the concerned subject matter.

Take, for example, the correlation coefficient derived from the MPT. It is such a “common knowledge” that people without any form of formal investment education could recite, word for word, why when choosing an asset to add to the existing portfolio, one should consider adding an asset that has low correlation with the existing portfolio. But how many of them know the assumptions needed to make such a recommendation? Yet most financial advisors and undergraduate students in finance would also make such recommendation because that is where most of the undergraduate investment courses’ coverage on MPT ends.

In this note, I demonstrate how easy it is, by using actual data, to show the inconsistent results obtained by MPT when we use the parameters estimated with data from one period to make asset allocation decision for the next period. I will compare the performances of minimum variance portfolio with historical data and the forecasted portfolio against a naïve investor.

II: The Minimum Variance Portfolio of Two Risky Assets

The key element in MPT is the existence of an efficient frontier that contains a dominant portfolio and a portfolio of assets that has a minimum variance among all other possible asset combinations. Effi-

cient asset allocation decision should be made using the risk and return information given by the dominant portfolio. The formation of the dominant portfolio depends on the correlation coefficients among all the assets within the asset universe.

One of the biggest hurdles in MPT is that the actual optimal portfolio doesn't really exist. Thus, neither the portfolio risk nor the return can be observed and must be estimated with sample observations. Recent literature suggests that the estimation of return is subject to large estimation errors (Roll, 1992). The estimation error for the variance is considerably smaller. For this reason, researchers and practitioners have focused on the minimum variance portfolio, which relies only on the second moments of samples (DeMiguel and Nogales, 2007). Once the minimum variance portfolio is obtained, the efficient frontier can be constructed, and the Capital Market Line (CML) can be derived with the risk-free asset return. Once the efficient frontier is known, an investor can make an asset allocation decision that meets the risk and return he/she desired efficiently. Thus, a successful efficient asset allocation hinges on forecasting performance of the parameters, such as the correlation coefficient, embedded in MPT.

It is easier to understand the MPT with a two-security portfolio. Most popular investment textbooks (see Bodie et al., 2009; Jordan and Miller, 2009; Strong, 2009; Reilly and Brown, 2009) present very similar examples. We can think of these two securities as combinations of a bond fund and an equity fund, or a mixture of two equity funds (or simply as two equities). The derivation of the minimum variance portfolio is available in all investment textbooks and is easily accessible for anyone interested in investing on their own. The following derivation is taken from one of the most common investment textbooks by Bodie et al. (2009).

Consider an individual investor who is considering how much he/she would invest in two securities, A and B. Let the portion of the fund invested in security A as w_a and the remainder of the fund be invested in security B (w_b). The expected return for this portfolio is:

$$E(r_p) = w_a E(r_a) + w_b E(r_b) \quad (1)$$

Variance of this two-security portfolio is:

$$\sigma_p^2 = w_a^2 \sigma_a^2 + w_b^2 \sigma_b^2 + 2w_a w_b \rho_{ab} \sigma_a \sigma_b \quad (2)$$

where σ_p^2 is the portfolio variance and σ_a^2 and σ_b^2 represent the return variance of securities A and B, respectively, and ρ_{ab} is the return correlation coefficient between securities A and B (value of ρ_{ab} is between -1 and 1). Since the sum of the weight in each asset must add up to 1,

that is, $w_a - w_b = 1$, thus, w_b equals $(1 - w_a)$. Substituting this expression for w_b gives us:

$$\sigma_p^2 = w_a^2 \sigma_a^2 + (1 - w_a)^2 \sigma_b^2 + 2w_a(1 - w_a)\rho_{ab}\sigma_a\sigma_b \quad (3)$$

Taking the derivative of the equation for σ_p^2 , setting it equal to zero, and solving for w_a , we get

$$w_a = \frac{\sigma_b^2 - \sigma_a\sigma_b\rho_{ab}}{\sigma_a^2 + \sigma_b^2 - 2\sigma_a\sigma_b\rho_{ab}} \quad (4)$$

This is the portion of the portfolio you should invest in asset A if you want to achieve a minimum variance portfolio.

Covariance can be computed from the correlation coefficient

$$\text{Cov}(r_a, r_b) = \rho_{ab}\sigma_a\sigma_b \quad (5)$$

Therefore,

$$\sigma_p^2 = w_a^2 \sigma_a^2 + w_b^2 \sigma_b^2 + 2w_a w_b \text{Cov}(r_a, r_b) \quad (6)$$

Other things being equal, portfolio variance is higher when ρ_{ab} is higher. In the case of perfect positive correlation, $\rho_{ab} = 1$, the righthand side of the above equation is a perfect square and simplifies to

$$\sigma_p^2 = (w_a \sigma_a + w_b \sigma_b)^2 \quad (7)$$

Or

$$\sigma_p = w_a \sigma_a + w_b \sigma_b \quad (8)$$

When the returns of two securities are perfectly correlated, there is no diversification benefit (i.e., the portfolio variance cannot be reduced by adding a security that is perfectly correlated with the existing portfolio). For all value of ρ_{ab} below 1, adding security B to the portfolio can reduce the overall portfolio variance.

Once the minimum variance portfolio is obtained, we can change the value of w_a (thus w_b) to obtain the efficient frontier. MPT suggests that, with the existence of a risk-free asset, an investor can determine an optimal portfolio by finding the tangency point of an imaginary line between the risk-free return and the efficient frontier. Therefore, the existence of an efficient frontier is a necessary condition for optimal portfolio construction.

The status quo in practice is to find two assets that have the lowest correlation coefficient with each other. A typical example in any investment textbook involves the historical returns, the standard variations, and a correlation coefficient between two assets. The correlation coefficient is allowed to change to form many different shapes for the efficient frontier. It is not uncommon for a textbook to include an Excel program whereby students can input those data and see how the efficient frontier changes with different values of the correlation coefficient. But the minimum variance portfolio indicated by equation (4) requires that the coefficient correlation be stable if the correct weight to be invested in each asset is fixed over a given period of time. If the correlation coefficient is time-varying, then the efficient frontier would also be time-varying. Thus, a static application of the MPT could lead to inefficient allocation of assets.

III. Methodology and Data

To demonstrate the potential problems associated with applying the MPT literally, I use actual data from the U.S. stock markets to construct the efficient portfolio frontier in one period. Then, I use the correlation coefficient obtained from the historical data of the first period and use all other parameters estimated from the second period historical data and reconstruct the efficient frontier. If the correlation coefficient is stable, which is the implied assumption when we recommend investors to use correlation coefficient as a criterion for asset allocation (since we can only observe this parameter from historical data), the efficient frontier of the forecast period should look very similar to the first period. If the forecast efficient frontier looks significantly different, then we must conclude that the correlation coefficient is not stable (a known fact). Thus, making the recommendation that one should include an asset that has low correlation coefficient with the existing portfolio is false. The correct recommendation should be that one should pick an asset that has low forecasted correlation coefficient (something not observable) with the existing portfolio.

Another problematic assumption of the MPT is that there are no transaction costs involved in the buying and selling of stocks. In practice, the transaction costs are very high for individuals who buy and sell limited shares of stocks in many companies. Therefore, it is not practical to assume that most individuals can buy individual stocks to construct their optimal portfolio. In academic research, it is common to assume that a market index is a good proxy for the market portfolio. With the development of Exchange Traded Funds (ETFs), it is also possible that individuals can buy and sell the market index at minimal

transaction costs. Therefore, the transaction cost issue can be minimized with the use of market indexes.¹ In this study, we assume that an individual will allocate funds among three stock market indexes in the U.S.: the Dow Jones Industrial Index (Dow), the S&P 500 Index, and the NASDAQ Index. These are the most well-known indexes and have many mutual funds and ETFs that would allow investors to diversify at a relatively lower transaction cost.

The Dow consists of the 30 largest companies in their representative industries. Twenty-eight of the 30 companies in the Dow are traded on the New York Stock Exchange (NYSE) (prior to 2006), while Intel and Microsoft are traded in the NASDAQ. The Dow is a less than ideal candidate for a market proxy because the index represents only a small sample of the over 8,000 publicly traded companies. Furthermore, it is a price-weighted index that is biased towards companies that have higher stock prices. Nonetheless, it is the most well-known index, and most people do consider the Dow as a market indicator.

I collected the value for all three indexes from 1996 to 2005 and divided the data into two periods. I used the data from the first 5 years to estimate the parameters needed to construct the minimum variance portfolio and efficient frontier. I also constructed the minimum variance portfolio and calculated the average return of the minimum variance portfolio. Next, I used the same parameters obtained from the first period data and applied them to the second period data to see whether it produces a similar shape for the efficient frontier. Furthermore, I assume that there a naïve investor who divides his/her fund equally among the indexes and compare the results with the values obtained from minimum variance portfolio. For robustness tests, I also use data from 1987 to 1996 for NASDAQ and S&P500.

IV. Results

Figures 1–4 display the results from pair-wise comparison between Dow-NASDAQ and S&P-NASDAQ. The results for the first period for both pairs (Figures 1 and 3) are exactly what is shown in all textbook examples. In fact, you can use any pair of stock in different industries and you could come up with a very similar efficient frontier. The MPT works remarkably well with historical data. Perhaps that is why very little attention has been paid to verification of the theory's forecasting ability.

¹I also conducted the tests with individual stocks. The results show an even higher degree of error. Thus, not only it is impractical from a transaction cost standpoint to use individual asset, it actually produces far inferior results.

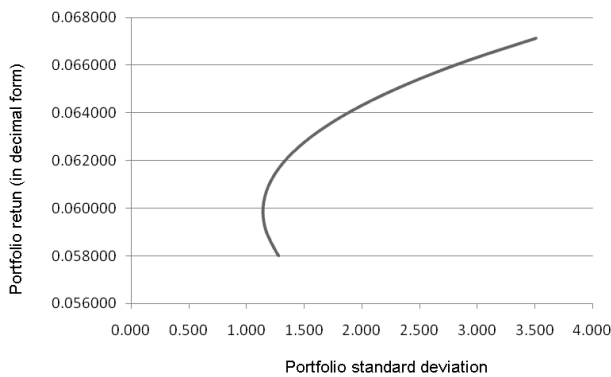


Figure 1. Efficiency frontier for Dow Jones Industrial Average and NASDAQ for Jan 2, 1996, to Dec 30, 2000.

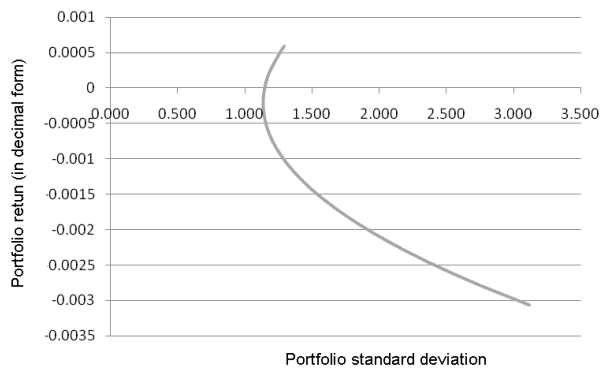


Figure 2. Efficiency frontier for Dow Jones Industrial Average and NASDAQ for Jan 2, 2001, to Dec 30, 2005.

The results for the second period (Figures 2 and 4), with the correlation value obtained from using the first period data, are completely different from those from the first period. This means that if an individual investor were to calculate the parameters needed to construct an efficient frontier using data from a particular period, the result could not be used for the second period. Therefore, the investor must recalcu-

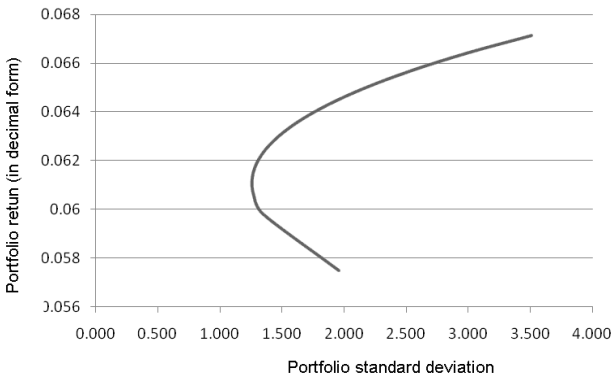


Figure 3. Efficiency frontier for S&P 500 and NASDAQ for Jan 2, 1996, to Dec 30, 2000.

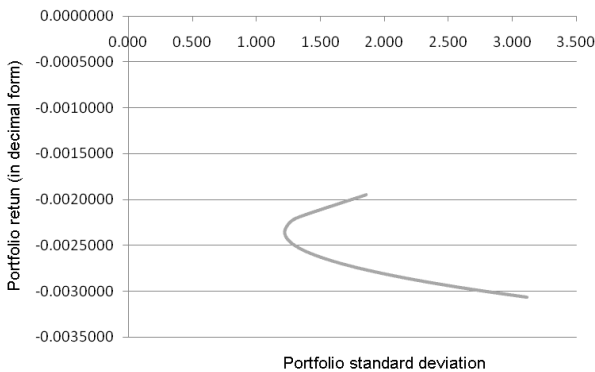


Figure 4. Efficiency frontier for S&P 500 and NASDAQ for Jan 2, 2001, to Dec 30, 2005.

late those parameters and rebalance the portfolio to keep the portfolio efficient. This would impose very high transaction costs on the investor. From a practical standpoint, this is not an option available to the average investor.

To check whether the result might be affected by the period chosen, I used data from post-1987 crash to 10 years after the crash. The result is shown in Figures 5 and 6. If anything, the result is even worse

than the result obtained from the first data set. Therefore, we can conclude that it is impossible for the investor to construct an optimal portfolio based on static data analysis.

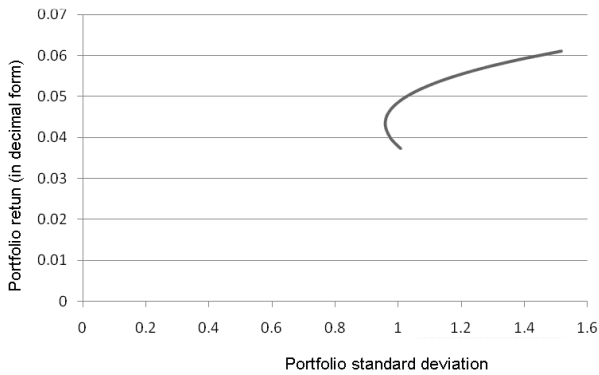


Figure 5. Efficiency frontier for Dow Jones Industrial Average and NASDAQ for Nov 1, 1987, to Oct 31, 1992.

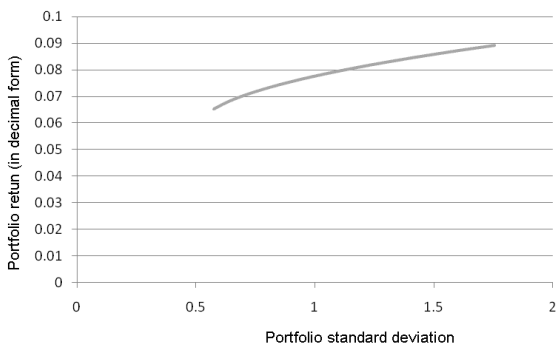


Figure 6. Efficiency frontier for Dow Jones Industrial Average and NASDAQ for Nov 1, 1992, to Oct 31, 1997.

What about the performance of the performance of the minimum variance portfolio? Kempf and Memmel (2003) argued that the global minimum variance portfolio often yields better out of sample results.

Could investors achieve better risk–return trade off by utilizing the parameters estimated from portfolio theory? To answer this question, I constructed a naïve portfolio by assuming that an investor will allocate 50% of his/her asset in either the S&P or the DOW and the other 50% in NASDAQ. The result is reported in Tables 1 and 2.

Table 1: Return, Standard Deviation, and Coefficient of Variation for the Minimum Variance Portfolio				
Portfolio	DOW-NASDAQ1	DOW-NASDAQ2	SP-NASDAQ1	SP-NASDAQ2
Return	0.058	0.0005	0.0598	-0.0022
Stdev	1.1120	1.1273	1.1618	1.3096
CV	19.1724	2254.6	19.428	-595.273

Table 2: Return, Standard Deviation, and Coefficient of Variation for the Naïve Portfolio				
Portfolio	DOW-NASDAQ1	DOW-NASDAQ2	SP-NASDAQ1	SP-NASDAQ2
Return	0.063	-0.0013	0.0636	-0.0026
Stdev	1.433	1.3533	1.4987	1.4029
CV	22.746	-1041	23.5645	-539.577

To compare the relative performance, we calculated the coefficient of variation (CV) for the DOW-NASDAQ and S&P-NASDAQ pairs for both the minimum variance portfolio and the naïve portfolio. The CV is defined as:

$$CV = \frac{\sigma}{\mu}$$

A lower value for CV suggests a higher risk–reward ratio. That is to say, for a given level of return, the risk associated with a portfolio with lower CV is lower than the risk associated with a portfolio with higher CV. In this regard, the minimum variance portfolio performs better than the naïve portfolio by a wide margin during the 1996 to 2005 period. That is, the risk–reward ratio for an investor whose sole concern is the lowest risk indicates the minimum variance portfolio works well. From a practical point of view, investors who are more interested in minimizing risk would probably be better off (or prefer) investing in the risk-free asset.

To verify the robustness of the results, I utilized the data from a different period. The 10-year data chosen included two of the largest

stock market crashes in history, which could distort the outcome. The last extended period in which there was no major correction in index value was from November 1987 to September 1997. Thus, I used data from that period to conduct the robustness test. I chose only to use the DOW-NASDAQ pairing because the robustness test should hold for any sample. The result, reported in Table 3, is drastically different than the results reported in Tables 1 and 2. In Table 3, the naïve portfolio performs better in terms of risk-adjusted return than the minimum variance portfolio.

Table 3: Return, Standard Deviation, and Coefficient of Variation, 1987 to 1997

Portfolio	DOW-NASDAQ1-M	DOW-NASDAQ2-M	DOW-NASDAQ1-N	DOW-NASDAQ2-N
Return	0.0373	0.050	0.0892	0.0781
Stdev	1.0035	1.0682	1.3247	0.932
CV	26.9035	21.364	14.851	11.9334

Therefore, we cannot conclude that, within the data range used in this study, a careful analysis of the static correlation relationships of the U.S. stock markets would provide any benefit to the investor, when compared with what can be achieved by a naïve investor who simply divides his/her stock market investment fund equally.

V. Conclusion

In this note, I demonstrated the problems with a literal interpretation of the MPT by using data from the U.S. stock markets from 1987 to 2005. I showed that it is only possible to construct an efficient frontier by static analysis of the stock market data with historical data. The forecast performance of the same parameters when used statically is mixed at best. Therefore, we cannot conclude that the correlation coefficient or the minimum variance portfolio derived from the MPT is of any use to an average investor. Some may argue that revisions and alternatives to the MPT that take into account uncertainty and continuous time adjustment have been proposed (see, e.g., Merton, 1969, and Samuelson, 1969); however, the practical application of these alternatives is far more difficult and costly, if not impossible.

The purpose of this paper is not to point out the many practical issues of the MPT. There have been numerous theoretical and applied studies that deal with this issue. Rather, the purpose of this paper is to demonstrate how easy it is, pedagogically, to show the poor forecasting

performance of a well-established theory, counter to what is commonly shown in textbooks. Most students who take investment courses could benefit from a more balanced approach to teaching MPT. MPT is a starting point for analysis. But it should not be the end.

References

Bodie, Zvi, Kane, Alex, and Marcus, Alan, 2009, *Investments*, 8th Edition, McGraw-Hill Irwin, New York, New York, USA

DeMiguel, Victor and Nogales, Francisco J., 2007, Portfolio Selection With Robust Estimation (August 2007). Available at SSRN: <http://ssrn.com/abstract=911596>

Jordan, Bradford and Miller, Thomas, 2009, *Fundamentals of Investments*, 5th Edition, McGraw-Hill Irwin, New York, New York, USA

Kempf, Alexander and Memmel, Christoph, 2003, On the Estimation of the Global Minimum Variance Portfolio (February 28, 2003). Available at SSRN: <http://ssrn.com/abstract=385760>

Markowitz, Harry M. 1952. Portfolio Selection, *Journal of Finance*, 7, 77–91

Merton, R.C., 1969, Lifetime portfolio selection under uncertainty: the continuous-time case, *Review of Economics and Statistics*, 51, 247–257

Reilly, Frank, and Brown, Keith, 2009, *Investment Analysis and Portfolio Management*, 9th Edition, South-Western Cengage Learning, Mason, Ohio, USA

Roll, R., 1992, A mean/variance analysis of tracking error. *Journal of Portfolio Management*, 18, 13–22

Samuelson, P., 1969, Portfolio selection by dynamic stochastic programming, *Review of Economics and Statistics*, 51, 239–246

Strong, Robert A., 2009, *Portfolio Construction, Management and Protection*, 5th Edition, South-Western Cengage Learning, Mason, Ohio, USA

An Analytical View of China as a Software Outsourcing Outlet

Taowen Le

Weber State University

Wayne Huang

Ohio University

Jin Zhang

University of Wisconsin Milwaukee

Abstract

While China is recognized as an important software outsourcing outlet today, its growth as a software outsourcing outlet is not without struggle. This article examines the early-day growth of China as a software outsourcing outlet, discusses hindering factors that once constrained China's potential as a software outsourcing service provider, and identifies various enabling factors that contributed to China's growing potential as a software outsourcing service provider. It analyzes the potential impact China's growing capabilities as a software outsourcing outlet may have internationally and identifies what China would need to do to become a greater contributor to the world's software out-

sourcing service industry. While the study focuses on China, its analysis and suggestions can be valuable to other developing countries as well.

1. Introduction

During the last decade, while China played a leading role in providing manufacturing outsourcing services, India undoubtedly led the world's software outsourcing services. Some researchers compared China and India and concluded that it was in the field of transaction costs, not production costs, where China was unable to compete with India in the supply of information technology (IT) outsourcing (Qu and Brocklehurst 2003), while others asserted that it was market fragmentation or small company sizes that kept China's software industry from grabbing a larger share of the global software-outsourcing market (Filippo and Hou 2005; Frauenheim 2005).

Researchers who compared the relative strengths of China's software industry with that of India identified a larger domestic software market, a larger amount of foreign direct investment, a better information and telecommunication infrastructure, and a more stable political and social environment. They further identified China's relative weaknesses as having poorer quality control processes and poorer communication skills. They subsequently suggested that China should speed up the infrastructure construction, build a consistent legal system, increase investment in general education and professional training, improve quality and process control, build correct channels to the global outsourcing market, and cooperate with India during competition (Hu et al. 2008).

China faces formidable challenges in having a lack of workers who could manage international projects and have strong English-language skills and having relatively poor protection of intellectual property rights (Benni and Peng 2008; Wang and Tian 2009). These researchers also analyzed certain unique strengths China possessed such as having two million Japanese and Korean speakers to help the nation's lead in the North Asian market for the "near-shoring" of IT services (Benni and Peng 2008).

After two decades of endeavors since the late 1980s, China's software industry has long grown from its infancy. As an emerging software outsourcing outlet or service provider, China has attracted attention from around the world during the last decade (Ji et al. 2008; Watson 2008; Jang et al. 2010).

Jones (2009) analyzed China's software outsourcing market from the perspective of international outsourcing companies and presented to them opportunities and challenges in the China market, while Jiang (2010) studied particular regions of China and suggested new growth points of outsourcing services in those regions.

Sarma (2005) attributed China's success in garnering investments in the IT services projects as well as for research and development to China's well-educated and cheap labor force. Einhorn (2009) felt that India's Satyam Mess helped China to win international clients.

Some researchers believe that the development of China's outsourcing service capabilities relied on support from both central and local authorities of China while concluding that despite the boom in China's service outsourcing, it would be difficult for China to catch up with India, the giant in the international outsourcing industry (Wang and Tian 2009).

Speculation developed that China would be a major threat to India's supremacy in software outsourcing services (Prasad 2008; Kathawala and Heeren 2009), while others believed China and India would codominate the world's software outsourcing market (Karthikeyan 2009).

Such speculations or beliefs are not without foundation. By 2010, China's software industry was no longer constrained within the lower end of software market (the market of simple or uncomplicated application software development projects); instead, it had successfully obtained a significant share of the middle and higher end of the market (the market of moderately complicated and highly complicated application software development projects) (Ma 2010). A Chinese government report states that during the decade from 2000 to 2009, the total revenue of China's software industry achieved a growth of 20 times, while the total export of China's software industry achieved a growth of 50 times. During the first seven months of 2010 alone, the total revenue of China's software industry reached 723.1 billion Chinese Yuan (¥), a 29% increase over the first seven months of 2009 (Revenue 2010).

While the world marvels at the rapid growth in capabilities and potential of China as a software outsourcing outlet, few understand its struggle in early days. On the other hand, the rapid growth of China's software outsourcing service might be perceived as a potential threat to India because of the lack of understanding of the positive international impact of China's growing capabilities as a software outsourcing outlet. Of course, although China's software industry has grown many folds since its early days, there is still much China needs to do to become a greater contributor to the world's software outsourcing service industry.

An extensive literature review has revealed that few researchers have taken the perspective of China with a growing potential as a software outsourcing outlet in analyzing China's outsourcing service market, few have studied the potential impact China's growing capabilities as a software outsourcing service provide might have internationally, and few, if any, have studied the China's outsourcing service market with a purpose or hope to offer insights to other developing countries.

This paper examines the early-day growth of China as a software outsourcing outlet, discusses hindering factors that once constrained China's potential as a software outsourcing service provider, and identifies various enabling factors that contributed to China's growing potential as a software outsourcing service provider. It analyzes the potential impact China's growing capabilities as a software outsourcing outlet may have internationally and identifies what China would need to do to become a greater contributor to the world's software outsourcing service industry. It is the hope of the authors that while the study focuses on China, its analysis and suggestions will provide valuable references and insights to other developing countries as well.

2. Early Days of China Software Industry

2.1. From Zero to a Tier 2 Country

In the early 1980s, China's government defined its strategic principle: "Economic construction must depend on science and technology, while work in these fields must be geared to the needs of economic construction" (Feng 1993). China's computer industry subsequently emerged.

For many years, however, China practically had only a computer hardware industry. In the late 1980s, overseas software companies began to enter China in search of market expansion, which began to draw the attention of China's IT professionals to software development.

During the following years, China's software development efforts dramatically increased. By 1995, Beyond Soft, a Beijing-based company, became the first software company in China to provide software outsourcing services (Dickinson 2010).

In June 2000, China's State Council issued the famous 18th Act—Policies Encouraging the Development of Software Industry and Integrated Circuits Industry—to further promote the growth of the software industry in China (State Council of China 2000). The issuance of the 18th Act established the foundation of the fast growth of China's software industry for the following decade. During the first six years, the total domestic and export sales of the software industry achieved an annual growth of over 30%. Table 1 shows the total software sales of

the country from 2000 to 2005 (Song 2004; Department of Information Industry of Liaoning Province, China 2006).

Table 1: Total Sales of China's Software Industry from 2000 to 2005	
Year	Total Sales (Unit: billion Chinese Yuan ¥)
2000	59.3
2001	79.6
2002	110.0
2003	160.0
2004	240.4
2005	360.6

Source: www.chinabyte.com/homepage/219001834121986048/20040323/1779910.shtml and Department of Information Industry of Liaoning Province, China

Meanwhile, China achieved a revenue of US \$633 million (or ¥5.064 billion) in software outsourcing services in 2004 and US \$920 million (or ¥7.36 billion Chinese Yuan) in 2005 (NeuSoft 2006).

In comparison with other countries, China was classified as a tier 2 country in the world map of software industry (Carmel, 2003), which was probably an accurate classification.

2.2. Hindering Factors that Once Constrained the Growth of China's Software Industry and Its Potential as a Software Outsourcing Outlet

For many years, China's software industry and its capability as a potential software outsourcing service provider were constrained by a few notable hindering factors.

2.2.1. Language Barrier

Unlike English or other western languages, which are typically based on alphabets or letters, the Chinese language is based on square-shaped characters. To people growing up in a character-based language environment, English is particularly difficult to master. Although the Chinese education system integrated English into its curriculum, English-speaking skills were still very lacking among the Chinese people (Benni and Peng 2008). It was estimated that in 2005, only about 0.77 percent of people in China could speak English (Kathawala and Heeren 2009).

Furthermore, ever since programming languages have existed, they have been English oriented. Developing software requires developers not only to understand English, but also to be able to communicate with international business partners including outsourcing companies. While general communication might get by with broken English, a failure to properly understand project specifications or documentation riddled with inappropriate English would generate serious consequences. The low percentage of people who were English literate in China combined with the fact that most English-literate people in China were English majors instead of computer science majors greatly constrained China's ability to expand into an English-dominant software outsourcing market. As a result, for many years, lack of English knowledge was an impediment to the Chinese software industry's expansion into English-speaking markets (Qu and Brocklehurst 2003; Zhang 2005).

2.2.2. Distrust in China's Government System

For decades, China's government systems remained a mystery to most Western people. Because the ruling party in China is the Communist Party, most Western people viewed China as a communist country instead of what it really was—a socialist country.

Prior to the 1980s, China's economy was essentially a planned economy controlled by the country's central government. In such an economic system, decision-making processes were typically not very transparent to the public, and business practices were governed more by government regulations than by a lasting legal system based on constitutional laws.

In the early 1980s, China adopted the open-door policy. Gradually, China became one of the most entrepreneurial countries in the world (Watson 2008). Because of the historical perceptions of China's ideological system and a lack of understanding of China's modern economic system, however, most Western people distrusted China's government system as a whole. Such distrust caused much hesitation and fear in the minds of Western business executives as they debated whether to outsource to China.

2.2.3. Lack of Confidence in Intellectual Property Protection

China was sometimes criticized for a lack of protection of intellectual property (Filippo and Hou 2005). While the criticisms were understandable, critics often failed to understand the historical root of the issue and the efforts China made to alter the situation.

Unlike Western developed countries, which had long practiced copyright and patent laws, China did not have to deal with intellectual property protection before the 1980s. Traditionally in China, intellectual property was typically owned by the national government and was available to everyone in the country free of charge.

Since privatization was introduced in China in the late 1970s, China has increased efforts by manifolds in educating its vast population to abandon the tradition and to respect other people's intellectual property; however, educating a population of 1.3 billion took time. Although China had enacted laws and established dedicated intellectual property protection bureaus throughout the country to promote the following of those laws, the speed of progress was not fast enough for most Western business executives who demanded a perfect match with Western countries.

One of the consequences was a lack of confidence by foreign businesses in China's intellectual property protection mechanism. As a result, many software development projects that could have gone to China landed elsewhere.

2.2.4. Lack of Competitiveness in International Competition

In the late 1980s, while Chinese professionals were still speculating about China's software market, overseas companies quickly occupied its various segments. Most of these overseas companies were well-established organizations with skilled employees, proven products, and many years of successful software development experience. Once they seized China's software market, they quickly fortified themselves by luring the most talented of China's developers with attractive salaries. Utilizing their management skills, existing products, market experience, and abundant resources, they soon raised the thresholds of market entry for local Chinese companies, their potential competitors. On the other hand, most of the newly founded local software companies did not possess the needed management skills, development capabilities or experience, and financial resources to compete with overseas companies.

A 2001 survey showed that China had over 10,000 software companies; however, only about 5700 of them had independent research and development capabilities (see Figure 1).

Compounded by a lack of experience and lack of financial resources to attract the top-quality researchers and developers, many local companies lacked creativity capability (Chen and Hu 2002). Out of the 5700 software companies that existed in 2001 and had some research and development capabilities, over 70% had a total number of

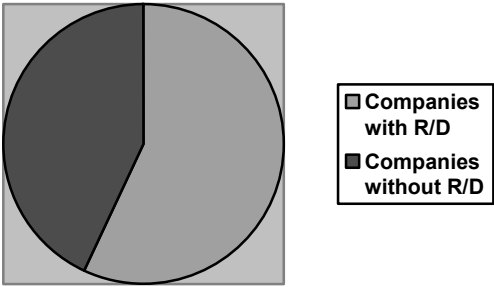


Figure 1. Research and development capabilities of Chinese companies
(Source: China Software Industry Association)

staff less than 50, and only 50 of them had a staff size of 1000 or more (see Figure 2) (Chen and Hu 2002).

Taking into consideration the fact that many software companies did not really develop software and the fact that many software companies also engaged in other types of business, such as retail, hardware assembly, realty management, and even hotel management, the true average size of software companies or software development groups was actually much smaller. Such a small size greatly constrained the strength of China’s software companies in global competitions.

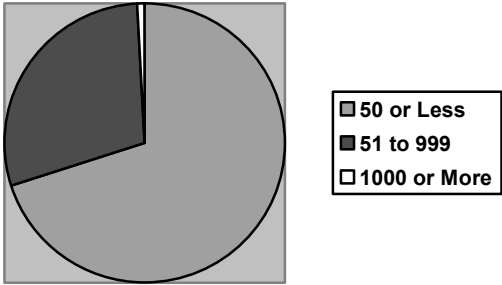


Figure 2. Company Size (Source: China Software Industry Association)

2.2.5. Lack of Investment Interest from Local Business Investors

Software development is resource demanding. On the one hand, it requires not only a large amount of initial investment in hardware and software, but also a large amount of ongoing investment because highly educated developers must be adequately compensated and other operating costs must be covered.

On the other hand, the financially “capable” investors in China typically did not become wealthy developing computer software. Realty, hotel, hardware manufacturing, and many other areas of business in China seemed to offer more immediate and less risky returns. Because software development neither offered any product as tangible as buildings, lands, and machines nor promised any immediate financial yield, investment in the software industry was simply too great a risk for China’s average business investors who typically did not possess the proper understanding of and experience with software industry.

What happened to the IT industry in the U.S. in the early years of the 21st century added more reluctance to Chinese investors. As IT bubbles cracked one after another in the U.S., without a proper understanding of the IT industry, average business investors in China became more certain that the IT industry would not be worth their consideration.

Furthermore, a perceived lack of strong intellectual property protection mechanism in 1980s and 1990s also caused much hesitation in China’s business investors in fueling the country’s software industry. On one hand, software development was so resource demanding; on the other hand, its final products could be so easily stolen away. The potential risk was simply too high for average Chinese investors.

Consequently, the development of China’s software industry was greatly hindered by the lack of investment interest.

2.2.6. Lack of Awareness among the General Public

Although computer information systems were first utilized by China’s business organizations in the early 1990s, their usage was not widespread. Only few government-owned financial institutions had the privilege to use computer information systems.

During the next 15 years, computer prices dropped significantly and keyboard input with the Chinese characters became more user-friendly. As a result, computers found their ways into many ordinary business buildings and private households in cities; however, the agricultural population constituted more than 70% of China’s total population. Because of a lack of knowledge and perceived needs, farmers in China typically would not spend their financial resources on strange luxuries such as computers.

Even to most of China's computer owners and users in cities, computer software still remained a mystery. Computers were generally and simply viewed as means of word processing, news publishing, emailing, online chatting, and gaming, not as a business strategic asset and valuable life-enhancing resource. Most of the public in China were not brought to the awareness of the great potential computer information systems could offer them.

Such lack of awareness among the public hindered the public imagination and desire for computer automation and constrained the domestic demand for more computer software.

Fortunately for China, most of the above-mentioned hindering factors or barriers have partially dissolved during the last six years. While language barriers are unlikely to dissolve completely over any short period of time, with a newer and more English-literate generation of college students joining the labor force and with the returning of numerous Chinese students who have completed their education in English-speaking countries, the size of English language obstacle has been significantly reduced. During the last six years, China has greatly increased the level of transparency in its government structure, legislative processes, and administrative decision-making processes. The establishment and enforcement of business laws have been much better publicized. As a result, China has won a much higher level of trust from foreign companies, although a complete trust in any foreign government would be difficult to achieve. China has also greatly increased its strength in enforcing laws and regulations in the area of intellectual property protection. Many violations of copyright and patent laws have been openly and severely prosecuted. Although many Western governments and companies still have concerns, they do recognize the great progress China has made. As more foreign companies came into China and as more Chinese companies studied the international market, many Chinese software companies have realized their lack of competitiveness in international competition. As a result, international standards are now better followed, and more business consortiums have been formed. However, although some Chinese companies have grown in competitiveness in the international market through following international standards and increasing company sizes, the majority of China's software companies remain uncompetitive. As China has encouraged the growth of its software industry through various government initiatives and especially as more software companies have proven their profitability, China has seen a moderate increase in private investments in software development; however, the majority of business investors still seem reluctant about investing in China's software industry. As more and more medium-sized and small business organi-

zations in China have benefited from the application of computer information systems, the general public is now more aware of the potential capacity of computer programs. However, the growth of public awareness has occurred primarily in urban areas; much of the agricultural population still need to be educated of the potential contributions computer information systems could make to their jobs and lives.

3. Enabling Factors of China's Growing Potential as a Software Outsourcing Outlet

Software outsourcing services in China could be traced back to the mid 1990s. As foreign software companies in China sought to cultivate China's software market, they needed local software developers to help them localize their products, which led to the emergence of early-day software outsourcing service providers in China. Later, the government worked hard to support the efforts of developing China as a potential software outsourcing outlet, and certain policies were implemented for this very purpose, such as the Initiative of Software Export to Europe and America instituted by China's Ministry of Science and Technologies in 2004 (Yuan 2006).

Despite the aforementioned hindering factors that constrained the growth of China's software industry and its potential as a software outsourcing outlet, certain important enabling factors greatly promoted the growth of China's software industry and its potential as an international software outsourcing destination.

3.1. New Domestic Needs Generated by International Competition

As China began to enjoy its rights and privileges of being a member of the World Trade Organization, it also fully opened its door to international business competition. The ever-intensifying global competition environment drove many domestic companies to seek improvement to their traditional approaches to production, management, and market research. Subsequently, computer automation became an important consideration for many business organizations in China (IBM Global Services 2006; Wang and Guo 2010).

3.2. Unique Needs Resulting from Linguistic, Cultural, and Social Backgrounds

While English-oriented programming languages made it difficult for small Chinese IT companies to compete internationally, the Chi-

nese-oriented market made it difficult for small overseas software companies to compete in China's software market (US Department of Commerce/Commercial Service 2008). Software products developed outside China had some key problems with China's local market. The products were typically designed outside China in a business and cultural environment different from that of China. Despite the localization process, they were more of a translation than redesign. Functions useless to the China market were not removed, and functions necessary to local users were either unavailable or put together by tearing and pasting (China Academy of Science 1997; Su and Lavina 2009). In addition, those software products were often constrained by regulations of the producing countries when released outside those producing countries. Moreover, because of the high development costs in developed countries, those software products typically were priced too high for Chinese users.

Such unique circumstances not only provided native Chinese software companies many opportunities to grow but also accelerated their maturity as software product developers and outsourcing service providers.

3.3. A Seemingly Infinite Supply of Human Resource

Software industry is a labor-intensive industry that needs educated human resource. One of the valuable assets China possessed was its abundant human resources (Li 2011). China had over 1200 colleges and universities, and the vast majority of them offered a computer science or information systems major. Since 1995, China has established many joint software schools between universities and software companies (Webplus 2004; Guo 2006). With a national population of more than 1.3 billion and a fast-growing education system, China possessed not only the largest potential market of the world, but also the largest potential supply of educated personnel for the global software industry.

3.4. Rapid Growth in Experience and Skills

Since their pioneering days in China, foreign software companies have helped educate many local employees (Li 1999; Newell 1999). These companies came to China not only for the country's vast market potential but also for inexpensive labor and IT professionals. As native employees changed their jobs, they also brought with them the experience and skills obtained in those foreign companies.

During the past three decades, tens of thousands of Chinese students have chosen to study or work abroad. Many of them obtained both quality education and employment experience in developed coun-

tries, including practical software development experience. As the IT industry in the U.S. and other developed countries was stumbling in recent years, many Chinese students and IT professionals returned to China for employment opportunities (European Commission and Ministry of Education in China 2011). They understood the Western cultures, systems, and language environments as well as those of China. As they returned to China, they brought with them not only such understanding, but also the experience and skills obtained abroad. They functioned as a bridge between the western world and China, which benefited businesses both in the West and in China.

Most importantly, as the software industry gained national emphasis, IT programs in China's higher education system as well as dedicated software engineering schools throughout the nation tremendously increased their efforts in training needed software personnel. Subsequently, the collective experience and skills of Chinese software developers greatly and quickly increased over the years.

3.5. The Strengthening of China's Intellectual Property Protection Mechanism

To facilitate an adequate software development environment, China not only increased efforts by manifolds in educating its vast population to respect other people's intellectual property during the past 25 years, but also enacted numerous laws and regulations to protect intellectual properties, including the Ordinance of Computer Software Protection in the People's Republic of China. Issued by the State Department on June 4, 1991, this ordinance specifies software user rights and obligations, facilitates software application development, and sets forth the legal relationship between software developers/providers and software users (Lu et al. 2002). In 2004, China enacted even stricter laws on intellectual property rights. Penalties for defiance of the laws have been raised significantly since then (Kathawala and Heeren 2009). These laws and regulations clearly established legal responsibilities of software users in China, which greatly helped cultivate a healthier environment for the development of its software industry.

3.6. The Adoption of International Standards

The IT industry has many standards. To succeed as an international software outsourcing destination, a company must follow international standards. During the past two decades, China greatly promoted the following of international software standards among its domestic software companies. Software companies such as NeuSoft

vigorously enforced the global CMMI5 standard (Xu et al. 2003; NeuSoft 2008). As a result, the capabilities of these companies in providing international software services were greatly enhanced.

4. International Impact of China's Growing Potential as a Software Outsourcing Outlet

There is little doubt that China's potential as an international software outsourcing service provider is still on the rise. Perhaps accustomed to India's leading the international software outsourcing service market, some people (especially those in India) view China as a potential threat to India's supremacy in software outsourcing services (Prasad 2008; Kathawala and Heeren 2009). While such concerns were understandable and respectable, we should also recognize the positive impact China's growing capabilities as a software outsourcing outlet might have internationally.

4.1. More Choices Available to the World

To countries that outsource their software needs, China's stronger presence in the outsourcing service market would provide one additional choice. Japan, for example, has greatly benefited from China's outsourcing services. Cultural and language differences or similarities often call for the availability of more service options.

4.2. Improved Products and Services for Outsourcing Clients

It is an undisputed fact that competition drives improvement. When the market becomes a buyer's market, the outsourcing companies will likely receive higher-quality services at better prices. On the one hand, with China's increasing potential as a software outsourcing service provider, current dominating countries will inevitably seek new ways to improve their products and services while keeping the prices down to safeguard their market shares. On the other hand, to survive and thrive in the market, China would be forced to continuously improve its own products and services while maintaining competitive prices. The outcome will be improved products and services for all countries that outsource their software needs.

4.3. Greater Availability of World's Resources for Innovations in Other Fields

The software industry is a labor-intensive industry, yet labor costs continue to rise throughout developed countries. With China joining the rank of maturity in software services, the world's collective software service capacity will increase tremendously. More and more countries would be able to outsource their lower-end software needs to service providers at considerably lower costs, thereby making available their valuable financial and human resources for innovations in other fields such as biotechnologies.

4.4. Increased Cultural Exchange and Mutual Understanding

As nations of the world become increasingly interdependent, the world needs cultural exchange and mutual understanding more than ever in human history. As pointed out by past researchers, however, distance still matters in our supposedly borderless world. Distance particularly constrains the synching of tacit knowledge, informal information, and cultural values (Heeks et al. 2001). As a result, cultural mishaps still occur within the context of global software development (MacGregor et al. 2005).

The success of cross-border outsourcing requires the cultural adaptation of the bridgehead teams working in the client countries (Krishna et al. 2004). As service requestors in one country and service providers in another country work together on software projects, many opportunities emerge for both sides to exchange viewpoints and gain more in-depth understanding about each other and the environments each other is in (see Figure 3).

With one fifth of the world's population living in China, increased potential for China as an outsourcing service provider means increased business interactions with the rest of the world and increased cultural exchange and mutual understanding among peoples of the world.

4.5. Encouraging Example for Other Developing Countries

Bordering China, Vietnam is similar to China in political, economic, cultural, and education systems. With a population of about 84 million and inspired by China's economic policies and renovations during the past decades, Vietnam became another possible IT outsourcing

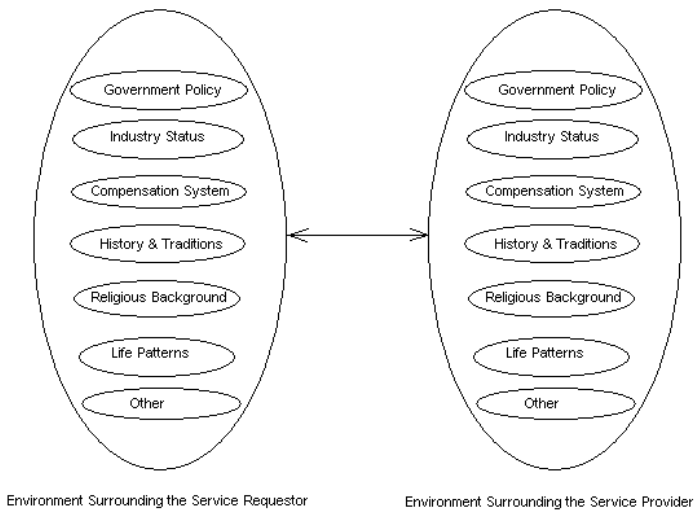


Figure 3. Cultural exchange and mutual understanding take place during software outsourcing

destination for western countries (Gallaughier and Stoller 2004; Yuan 2006). As China improves its potential as a software outsourcing outlet, it will undoubtedly become a valuable model and encouraging example for developing countries such as Vietnam.

5. What China Should Do to Become a Greater Contributor to the Global Software Outsourcing Service Industry

China's potential as an international software outsourcing outlet has been recognized not only by researchers but also by major companies like IBM, Oracle, and Hewlett-Packard (Watson 2008; Dickinson 2010). It is believed that if the IT industry develops as expected, China could capture software outsourcing opportunities worth \$56 billion by 2015 (Kathawala and Heeren 2009).

To become a greater contributor to the world's software outsourcing service industry, China should continue its efforts to overcome relative weaknesses and fortify relative strengths in its software industry.

5.1. Prepare for Increased Software Outsourcing Needs from Japan

Table 2 lists the major countries that outsource their software needs as well as the primary service providers that service these needs. It has been noted that Japan has the second largest software industry in the world and yet only 1% of its software needs have been outsourced, and that 70% of Japan's software outsourcing needs have been serviced by China (Yuan 2006). As a leading country in software development in the world, Japan was one of the earliest countries in the world to have deployed information systems in its business world; however, as the country becomes ready to replace or update their legacy systems, and with personnel costs in Japan being considerably higher than those in developing countries, more software projects can be anticipated to be outsourced. With geographical and language proximities between China and Japan, and with past success servicing Japan's software outsourcing needs, China would be a logical potential outlet for Japan's increased needs. China's active preparation would undoubtedly strengthen its candidacy as a service provider for the anticipated software outsourcing needs in Japan.

Table 2: Major Outsourcing Countries and Corresponding Service Providers

Major Service Requestors	Major Service Providers
U.S.	India
West European countries	Ireland
Japan	China

Source: CSIACW 2006 and Yuan 2006

5.2. Upgrade the Size of Software Companies

As established by the Software Engineering Institute, the Capability Maturity Model (CMM) measures the ability of software development companies to produce quality software within budget and on schedule (Saiedian and Kuzara 1995). The capability maturity level is one comprehensive index software outsourcing companies carefully examine in software service providers.

In well-cited research, Bharadwaj (2000) defined a company's IT capability as "its ability to mobilize and deploy IT-based resources in combination or co-present with other resources and capabilities." In his classification of the IT-based resources, human IT resources comprising the technical and managerial IT skills were ranked second.

Compared with software companies in India, the competitiveness of Chinese software companies in servicing large-scale software projects has been typically constrained by their relatively small company sizes. Table 3 compares the number of software service companies and the number of large software service companies in China and in India as of 2006. Although certain mergers have taken place during the past few years in China, compared with India, China still does not have many large IT companies (Einhorn 2009).

As larger companies tend to generate in customers a feeling of assurance of higher creditability and lower business risk, they will more likely succeed in winning larger and upper-end contracts from potential outsourcing clients. For example, India's top 10 companies service about 45% of the world's software outsourcing needs (Kathawala and Heeren 2009).

Table 3: Software Companies in China and India		
	China	India
No. of software companies	8000	3000
No. of software companies with at least 2000 employees	5	15

Source: CSIACW 2006

To become more capable of servicing larger software projects from other countries, China needs to continue upgrading the sizes of its software companies. Possible approaches include mergers and consortium formation.

Market economy has been practiced in modern China for only about 30 years, and merger is a relatively new concept; however, numerous mergers have already successfully taken place in other business areas in China. Through mergers, small groups of resources could be combined into major source of strength in global competition.

One alternative would be to form consortiums among software companies. Software companies within a given geographical location typically complement, rather than compete with, each other. The formation of a consortium with these companies allows them not only to share the responsibilities and risks collectively, but also to serve the market with their collective strength. As a result, the perceived risk in service requestors' outsourcing software projects to them would be significantly smaller, and their chance of winning the confidence of potential service requestors would be greatly improved.

5.3. *Continue Improving Quality of Products and Services*

The global software outsourcing service market is no different from other areas of business in that superior product and service quality sustains lasting business relationships. Software companies in China must constantly improve the quality of their software products and services to become more significant contributors in the global software outsourcing service industry.

To improve quality of products and services, software companies in China must follow widely-accepted international standards. As depicted in Figure 4, as of 2006, among the 30 largest software companies in China, only 6 achieved the level 4 or 5 of CMM Certification; in India, 100% of its 30 largest software companies made such achievement. More disturbing was the finding by a 2006 survey of the Chinese Software Industry Association that most of the Chinese software companies expressed a lack of interest in CMM standards (CSIACW 2006).

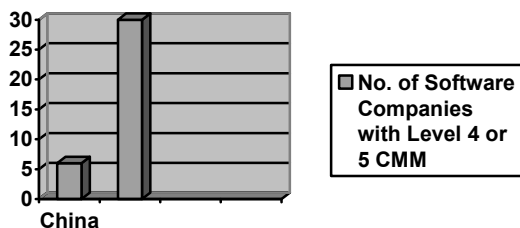


Figure 4. CMM Certification of 30 Largest Software Companies in China and India (Source: CSIACW 2006)

A lack of standards often indicates no quality assurance. As the IT industry is an industry of standards, widely accepted international standards must be followed if a software company or a country wishes to become a significant player in the international market.

To motivate Chinese software companies to follow widely accepted international standards, China must educate its university faculty and IT professionals about the value of the CMM standards to China's software industry and the importance of following those standards in their efforts to expand into the international market.

To improve quality of products and services, software companies in China must also maintain stable and quality human resources. It was reported that the annual turnover rate in China's software companies

reached as high as 20% (CSIACW 2006). With the relatively young age of China's software industry and with the emergence of so many software companies in the recent years, the seemingly abundant opportunities do present much attraction to software developers; however, companies could maintain relatively stable development forces by offering employees various incentives such as profit sharing, stock options, or dual-career in which employees are offered a non-management, technical advancement track with compensation packages identical or comparable with those of the management advancement track.

The IT industry is one of quick updates and frequent renovations, and the software industry is no exception. China's software companies have not paid much attention to ongoing training of their current employees. To maintain continuously qualified development forces and to assure high quality in products and services, software companies must invest in the ongoing professional development of their human resources.

To improve the quality of products and services, software companies in China should specialize instead of generalizing. Driven by a desire to diversify their revenue channels, many software companies in China engage in large varieties of software production. Consequences usually include weakened research and development forces and software products that lack international competitiveness. Unless a company has sufficient human resources to effectively thrive in multiple fields of software development, it should carefully select a field of specialization and establish itself in that field before it attempts to expand into other areas. Through specialization, a company will be able to improve the quality of its software products and services more easily.

5.4. Explore New Service Areas

While specialization would benefit individual companies, China as a country should expand its software services. As a country, China has been relatively weak in systems software development, and most of its software companies have chosen application software development. Therefore, China's software companies have not been very competitive in winning large contracts that usually include the integration of systems software. A possible remedy would be for larger companies that have sufficient human resources to dedicate special development forces, or to join efforts with other companies, to explore the sector of systems software, especially system software integration. Once they are experienced, these companies would be positioned to service larger and more complex software projects.

5.5. Strengthen the Enforcement of Intellectual Property Protection Laws and Regulations

During the past years, China has enacted various laws and regulations protecting intellectual properties. While these laws and regulations have clearly established legal responsibilities of software users in China, the enforcement of these laws has frequently become a target of criticism from other countries. Although many understand from China's historical background that intellectual property used to be owned by the national government and made available to everyone in the country, most countries expect a complete conformance of international standards from China now if China is to participate in international business. Therefore, China must on the one hand accelerate its efforts to educate its citizens, and on the other hand, strengthen the enforcement of the laws and regulations already established.

5.6. Understand business process and culture in the U.S. and the E.U.

With China's doors opened to the world only three decades ago, the typical U.S. and E.U. ways of conducting business are still quite unfamiliar to the majority of business people in China, yet most of today's software outsourcing needs originate from the U.S. and the E.U. countries. To improve their competence in meeting outsourcing needs from the western world, China's software service companies must understand the typical business processes and the culture specialties of the western countries.

5.7. Native Chinese Companies Must Know How to Win Trust from Clients

It is common knowledge that personal relationships play an important role in business in East Asian countries including China (Jang et al. 2010). Such reliance on personal relations tends to generate an element of unpredictability and create distrust in western clients as they typically lack such relationships in foreign countries such as China. To win client trust, Chinese software service companies must learn to base their business operations on well-accepted and well-followed international business standards. Furthermore, they should maintain good relationships with their clients even after the conclusion of their present service contracts.

6. Conclusion

One of the most important factors outsourcing companies consider when they select potential service providers is risk diversification. While countries such as India and Ireland have long established themselves as major service providers in the international software outsourcing market, more and more companies are willing to consider newly emerged service providers such as China. Given China's steady supply of human resources and rapid growth in experience and skills, China is expected to play a more important role in the global software outsourcing service industry.

As China increases its efforts in software outsourcing services, international competition in this field will inevitably intensify; however, the increased competition will bring to outsourcing companies or countries greater choices of business partners, improved products and services, and greater availability of resources for innovations in other fields. In addition, the world will enjoy increased cultural exchange and greater mutual understanding. To developing countries such as Vietnam that have political, economic, and education systems similar to those of China, China's success would undoubtedly serve as an encouraging example.

To become a greater contributor to the global software outsourcing service industry, China should prepare for the upcoming increase of software outsourcing needs from Japan; continue to upscale the sizes of its software companies by mergers or consortium formation; continue improving the quality of its software products and services by following widely accepted international standards, maintaining stable and quality human resources, encouraging individual companies to specialize in specific areas of software development, expanding service scope as a country; and toughen the enforcement of intellectual property protection laws and regulations already enacted. China's software companies need to better understand the business processes and business culture of American and European countries and win the trust of business people in these countries. Although this study focuses on China, its analysis and suggestions can be valuable to other developing countries as well.

References

1. Benni, E. and Peng, A. "China's Opportunity in Offshore Services," McKinsey Quarterly, No. 17, 2008.
2. Bharadwaj, A. S. "A Resource-Based Perspective on

Information Technology Capability and Firm Performance: An Empirical Investigation,” 2000. *MIS Quarterly*, Vol. 24, No. 1, 169-196.

3. Carmel, E. “Taxonomy of New Software Exporting Nations,” 2003. *Electronic Journal of Information Systems in Developing Countries*, Vol. 13, No. 2, 1-6.

4. Chen, C., and Hu, K. “An Overview of the Current State and Development Trends of China’s Software Industry,” 2002. Available at <http://www.e-works.net.cn/ewkArticles/Category117/Article5432.htm> (Accessed Sept. 6, 2004).

5. China Academy of Science. “The Design and Implementation of Kaisi ERP System,” 1997. Available at <http://code.google.com/p/erpbbs/downloads/list> (Accessed June 17, 2011).

6. CSIACW. “The Current State and Future Trend of China’s Software Outsourcing Service Market,” 2006. Available at <http://articles.e-works.net.cn/asp/article40485.htm> (Accessed Oct. 10, 2006).

7. Department of Information Industry of Liaoning Province, China. 2006. Staff presentation.

8. Dickinson, K. “Beijing Offers Capitalism, Technology,” *Phoenix Business Journal*, November 1, 2010. Available at <http://www.bizjournals.com/phoenix/blog/business/2010/11/beijing-offers-more-capitalism-for.html> (Accessed March 15, 2011).

9. Einhorn, B. “China Aims to Gain from Satyam Mess,” *Business Week*, January 14, 2009.

10. European Commission and Ministry of Education in China. “EU-China Student and Academic Staff Mobility: Present Situation and Future Developments: Joint study between the European Commission and the Ministry of Education in China,” April 2011. Available at http://ec.europa.eu/education/external-relation-programmes/doc/china/mobilitysum_en.pdf (Accessed June 20, 2011).

11. Feng, J. “Legend in Chinese Computer Industry,” *Beijing Review* (6), 1993, 25-29.

12. Filippo G. and Hou, J. "Can China Compete in IT Services?" McKinsey Quarterly, No. 1, 2005, 10-11.
13. Fraunheim, E. "Report: China's Outsourcing Industry Lags India's," CNET News, February 3, 2005. Available at http://news.cnet.com/Report-Chinas-outsourcing-industry-lags-Indias/2100-1011_3-5562791.html (Accessed March 17, 2011)
14. Gallagher, J., Stoller, G. "Software Outsourcing in Vietnam: a Case Study of a Locally Operating Pioneer," 2004. Electronic Journal of Information Systems in Developing Countries, Vol. 17, No. 1, 1-18.
15. Guo, J. "PacificNet Reports Record Fourth Quarter and 2005 Year-end Results," April 18, 2006. Available at <http://www.douban.com/group/topic/1068683> (Accessed June 20, 2011).
16. Heeks, R., Krishna S., Nicholson B., Sahay, S. "Synching or Sinking: Global Software Outsourcing Relationships," 2001. IEEE Software, Vol. 18, No. 2, 54-60.
17. Hu, H., Lin, Z. and Foster, W. "China's Software Industry – Current Status and Development Strategies," 2008. Available at <http://tradeinservices.mofcom.gov.cn/en/i/2008-09-03/54732.shtml> (Accessed June 21, 2011).
18. IBM Global Services. "Going Global: Prospects and Challenges for Chinese Companies on the World Stage," IBM Corporation, 2006, 1-16. Available at <http://www-935.ibm.com/services/us/imc/pdf/g510-6269-going-global.pdf> (Accessed June 16, 2011)
19. Jang, S., Lee, H., Ko, K. and Kim, D. "Evolution of Offshore Software Outsourcing in East Asian Countries: From Cost to Relation," Proceedings of the Pacific Asia Conference on Information Systems, Taipei, July 9-12, 2010, 1604-1609.
20. Ji, J., Li, J., Conradi, R., Liu, C., Ma, J. and Chen, W. "Some Lessons Learned in Conducting Software Engineering Surveys in China," Proceedings of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, Kaiserslautern, Germany, October 9-10, 2008, 168-177.
21. Jiang, R. "To Develop Outsourcing: New Growth Point of

Northeast China Using FDI,” 2010. *International Business Research*, Vol. 3, No. 1, 162-165.

22. Jones, W. “Outsourcing in China: Opportunities, Challenges, and Lessons Learned: Industry Insight,” 2009. *Strategic Outsourcing: an International Journal*, Vol. 2, No. 2, 187-203.

23. Karthikeyan, M. “Is China a Threat to Indian IT Outsourcing?” *ITonion*, January 15, 2009. Available at <http://www.itionion.com/2009/01/is-china-threat-to-indian-it.html> (Accessed March 17, 2011)

24. Kathawala, Y. and Heeren, C. “IT Outsourcing: China Grasps for the Lead,” *Graziadio Business Review*, Vol. 12, No. 3, 2009. Available at <http://gbr.pepperdine.edu/2010/08/it-outsourcing-china-grasps-for-the-lead/> (Accessed March 15, 2011)

25. Krishna, S., Sahay, S., Walsham G. “Managing Cross-Cultural Issues in Global Software Outsourcing,” 2004. *Communications of the ACM*, Vol. 47, No. 4, 62-66.

26. Li, J. “Avoiding the Middle-Income Trap,” *Beijing Review*, March 3, 2011. Available at http://www.bjreview.com.cn/quotes/txt/2011-03/02/content_336142.htm (Accessed June 20, 2011).

27. Li, S. “Management Development in International Companies in China,” *Education + Training*, Vol. 41, No. 6/7, 1999, 331-336.

28. Lu, W., Du, J., Zhang, J., Ma, F., and Le, T. “Internet Development in China,” *Journal of Information Science* (28:3), 2002, 207-203.

29. Ma, H. “Shell-Breaking of China Software Industry,” *China Internet Weekly*, No. 5, 2010. Available at <http://tech.sina.com.cn/it/2010-03-08/20483916575.shtml> (Accessed March 16, 2011)

30. MacGregor, E., Hsieh, Y., Kruchten, P. “Cultural Patterns in Software Process Mishaps: Incidents in Global Projects,” 2005. *ACM SIGSOFT Software Engineering Notes*, Vol. 30, No. 4, 1-5.

31. NeuSoft, “NeuSoft Software Outsourcing Services Rank No. 1 in China in 2005,” 2006. Available at <http://www.neusoft.com/cn/news/html/20060221/1792160421.html> (Accessed Oct. 10, 2006).

32. NeuSoft, "Neusoft Receives Top Ranking among All Offshore Software Outsourcing Providers in China in 2007," April 12, 2008. Available at <http://www.chnsourcing.com/article/oic/Informations/225520080412092107.html> (Accessed June 20, 2011).
33. Newell, S. "The transfer of Management Knowledge to China: Building Learning Communities Rather Than Translating Western Textbooks," *Education + Training*, Vol. 41, No. 6/7, 1999, 286-294.
34. Prasad, S. "China Is India's 'Only Possible Threat'," ZDNet, May 9, 2008. Available at: <http://www.zdnetasia.com/china-is-india-s-only-possible-threat-62041177.htm>
35. Qu, Z. and Brocklehurst, M. "What Will It Take for China to Become a Competitive Force in Offshore Outsourcing? An Analysis of the Role of Transaction Costs in Supplier Selection," *Journal of Information Technology*, No. 18, 2003, 53-67.
36. "Revenue of Our National Software Industry This Year Will Surpass 1200 Billion," September, 08, 2010. Available at <http://www.cnbeta.com/articles/121370.htm> (Accessed March 17, 2011)
37. Saiedian, H. and Kuzara, R. "SEI Capability Maturity Model's Impact on Contractors," 1995. *Computer*, Vol. 28, No. 1, 16-26.
38. Sarma, V. "IT Outsourcing: China versus India," *Frost & Sullivan Market Insight*, June 8, 2005. Available at <http://www.frost.com/prod/servlet/cif-econ-insight.pag?docid=39745707> (Accessed March 17, 2011)
39. Song, B. "China's Software Industry Surpassed India in Sales and Software Companies Totaled 8582," 2004. Available at <http://www.newer.cn/news/3526.shtml> (Accessed Sept. 6, 2004).
40. State Council of China. "Policies Encouraging the Development of Software Industry and IC Industry," 2000. Available at <http://www.chinaacc.com/new/63/67/81/2006/2/bi00720102260020-0.htm> (Accessed Oct. 16, 2006).
41. Su, N. and Lavina, N. "Internationalization Strategies of IT Vendors from Emerging Economies: The Case of China," 2009. Available at http://blogs.nyu.edu/blogs/jr197/japan/Su_and_Levina_N010.pdf (Accessed June 17, 2011).

42. U.S. Department of Commerce/Commercial Service. "China: Customs, Taxes and Documentation Requirements for IT Products and Service Imports," 2008. Available at <http://web.ita.doc.gov/ITI/itiHome.nsf/9b2cb14bda00318585256cc40068ca69/3aed69c22069b76185256d07006fe126?OpenDocument> (Accessed June 17, 2011).
43. Wang, F. and Guo, X. "Status and Trends of IT/IS Application in Small and Medium Manufacturing Enterprises in China," 2010. The 2010 International Conference on E-Business Intelligence, Atlantis Press, 194-202.
44. Wang, T. and Tian, G. "IT Outsourcing: China vs. India," Bangkok Post, March 21, 2009.
45. Watson, B. P. "Will China Change IT?" CIO Insight, August 27, 2008. Available at <http://www.cioinsight.com/c/a/Expert-Voices/Will-China-Change-IT> (Accessed March 17, 2011)
46. Webplus, Inc. "Victor Tong is Invited to be a Guest Professor at College of Software of Beihang University," January 8, 2004. Available at <http://usa.webplus.com/company/presstext.en.wml?id=139> (Accessed June 20, 2011).
47. Xu, Y., Lin, Z. and Foster, W. "Agile Methodology in CMM Framework: an Approach to Success for Software Companies in China," GITM 2003, Calgary, Canada, June 8-10, 2003.
48. Yuan, H. "China's Software Outsourcing Services Challenge India," 2006. Available at <http://news1.jrj.com.cn/news/2006-09-29/000001683950.html> (Accessed Feb. 27, 2007).
49. Zhang, J. "Special Report: Outsourcing to China," 2005. Available at <http://www.sourcingmag.com/content/c050803a.asp> (Accessed June 14, 2011).

The Impact of Workplace Wellness Programs on Decreasing Employee Obesity and Increasing Overall Health

Skyler Charles MacDonald and Jonathan H. Westover
Utah Valley University

Abstract

Obesity is a deadly disease that is spreading across America and creating an epidemic that needs to be stopped. The risks are very serious and have a great impact on the quality of the lives of those with obesity. The obese individual is not the only one who suffers the consequences of their obesity. The whole of society is being affected by this epidemic. The costs of obesity affect not only the individual but the company that employs him or her. A solution to obesity is through creating a corporate culture that is physically active where the company provides employee benefits that not only increase the well-being of the employees but also improve the quality of their lives. Corporate wellness programs benefit the whole of a business both financially and through the health of the employees. It is important to a company to improve and sustain the health of its employees. The implementation of wellness programs can influence the employees in living healthier lives. There needs to be a change. Employers are in the best position to make the largest impact on the greatest amount of people.

Americans are working in a more sedentary environment where there is minimal physical activity during the work day (Berry et al., 2006). Since there are increased efficiencies there is less need for physical labor, which is one reason people may be less fit. The workforce is becoming unhealthy through lack of exercise, bad eating habits, and change in lifestyle (Rippe and Hess, 1998). There is a need for the implementation of health promotion programs within businesses. A change in the behavior of employees is necessary to create a corporate culture that values personal well-being and promotes physical activity (Anderson et al., 2009).

We will be discussing the facts about obesity and the spread of this epidemic. We will show how corporate wellness programs benefit the whole of a business both financially and through the health of the employees. We will explore examples of businesses that have different health programs in place to improve the health of employees, faculty, students, or staff. The costs of obesity are felt by both the company and the employees. We will demonstrate how wellness programs are a solution to the rise in costs of obese and overweight employees.

Americans watch as their nation is plagued with an obesity epidemic affecting all citizens. There has been a dramatic increase in obesity among children, adolescents, and adults (Sturm, 2007). There may not be a quick fix to this problem, but with time we can develop new ways to combat this epidemic and create a society focused on healthy living. Rather than letting our nation get deeper and deeper into the problem, we should strive to provide a solution. It is our responsibility to come up with new and creative concepts that will eliminate obesity and help our nation progress.

The costs of obesity affect not only the individual but the company that employs him or her. According to the National Center of Health Statistics, 68% of Americans are either obese or overweight. In this paper, we will show how this is affecting the whole company from the extra health care costs the employers are paying to the productivity of the employee.

Through this research, we were able to see the effects of certain activities on employees and their companies. Some of these activities include physical exercise, recreation, healthy eating, and incentives in the workplace. The study of effective behavioral change is crucial to companies that are responsible for paying the rising costs of obese workers. Some of these costs include physical illness, mental distress, lower productivity, higher turnover, and increased healthcare spending. This study has included the evaluation of effective programs that some companies have been implementing. There needs to be a cost-effective way for health programs to grow and be successful. Health promotion

programs benefit both the individual and the company. A solution to obesity is through creating a corporate culture that is physically active where the company provides employee benefits that not only increase the well-being of the employee but improve the quality of his or her life.

Is Obesity a Problem?

Obesity is a pressing issue in America. Our nation's Surgeon General has stated in *The Surgeon General's Vision for a Healthy and Fit Nation 2010* that, "Today's epidemic of overweight and obesity threatens the historic progress we have made in increasing American's [sic] quality and years of healthy life. Two-thirds of adults and nearly one in three children are overweight or obese" (U.S. Department of Health and Human Services, 2010, p. 4). This epidemic is killing and destroying the citizens of this nation. We still have not found a solution for reducing obesity. Our country has an internal problem that has affected so many people. We should act now to change this. Hundreds of people die every day from this epidemic that is spreading across our country. The problem is not only affecting adults but also children. We are producing children whose life expectancy is not as long as our own (Yackobovitch-Gavan and Nagelberg, 2009). We have not found a solution because obesity has many sources.

Obesity is a very complex issue. It is not just about eating less and exercising more. Some of these complex issues deal with genetics, social factors, and stress. These are important pieces of the problem that need to be addressed. They may even be as important as diet. With the combination of factors affecting people in becoming obese, it is important to understand each of them. There are many risks for people who are overweight or obese. Overweight individuals have a high risk of developing cardiovascular disease and type II diabetes. Many types of cancer have been linked to excess body weight. The risks are very serious and have a great impact on the quality of the lives of those with obesity (Bianchini et al., 2002).

Americans seem to always be overly stressed and try to cram as much as they can into the day. One reaction to this stress is a craving for food (Hristova and Aloe, 2005). When a person is stressed the metabolism slows down and the body goes into a storage state where it operates as if it is under famine conditions, and it does not burn fat. Stress creates cortisol in the adrenal glands. The more stressed a person is, the chances of obesity increase. Cortisol tells the brain to eat and the stomach to store. That is the response that bodies have when in stressful situations (Hristova and Aloe, 2005).

In addition, the food we eat is creating havoc with our bodies. We are attracted to foods that contain sugar, fat, and salt. Scientists are able to see the damaging affects of sweets on our bodies. One interesting aspect of sugar is that it affects the same system of the brain that illegal drugs do. It has been shown on rats in the laboratory that “drinking a sugar solution on an intermittent schedule can promote sugar bingeing and cause signs of dependence while releasing dopamine repeatedly like a drug of abuse” (Avena et al., 2006, p. 813). Thus, sugar is known to induce a form of dependency.

The Centers for Disease Control and Prevention (CDC) says that, “Policy and environmental change initiatives that make healthy choices in nutrition and physical activity available, affordable, and easy will likely prove most effective in combating obesity” (www.cdc.gov). This may prove harder than they say it really is. We have created a system in which the least healthy calories in the supermarket are the cheapest and the healthy calories are the most expensive. If it was not so expensive to eat healthy, then more people might purchase healthier foods. With the abundance of cheap food and large portion sizes, America seems to be eating more and more every day. A child-size fast food meal is actually an adult-appropriate size that contains around 700 calories. Adult meals are, in fact, a family-appropriate size with 1700 calories. With food like this at the cost of mere dollars, most people will choose this over an expensive healthy meal. It may be hard to beat the lure of inexpensive food, but people can change their eating habits.

How do we, as Americans, change ourselves back to how we used to be? In earlier times, two thirds of the population was not obese. We did not have fast food restaurants on every corner. How do we encourage people to live a healthier lifestyle? We have to deal with this problem as a comprehensive society. There needs to be a motivator to get people moving in the right direction. Change can be made when everyone comes together and says that we are not getting what we need out of this system. We have to make it better ourselves.

The Centers for Disease Control and Prevention

The CDC is geared to helping America combat the spread of obesity. The Division of Nutrition, Physical Activity, and Obesity (DNPAO) is working to “reduce obesity and obesity-related conditions through state programs, technical assistance and training, leadership, surveillance and research, intervention development and evaluation, translation of practice-based and research findings, and partnership development” (www.cdc.gov). Even with all of this going on, obesity is still increasing and Americans are becoming heavier. It is not just about

getting healthy through eating right and exercising. It is a lifestyle change. Obesity will not be cured with a pill or some magical formula. Each person is different; therefore each case of obesity will be unique.

The CDC has created a wellness program called “CDC's LEAN Works—Leading Employees to Activity and Nutrition.” It is a free resource that includes programs and tips to help establish an effective wellness program in the workplace. It also includes a calculator to see how much obesity is costing a company and how much the establishment of different wellness programs will save the company. The CDC created this free resource in response to the many inquiries about which programs work in combating the obesity epidemic. Since companies do not publish information regarding their programs, “CDC visited selected businesses to identify promising worksite obesity prevention and control practices” (www.cdc.gov). The programs that the CDC recommends have been backed by science and evidence-based research. Through their website, any employer or employee can access the CDC guide to developing a wellness program for their company. The website will help guide the employer through the steps needed to implement a program that will be useful and cost effective. They even help with the planning, building, promoting, and assessing the program. It is important for every company to promote health throughout the workplace. With a program in place, employees can have the resources available to maintain a healthy lifestyle.

Obesity in the Workplace

Obesity affects the whole company because of the extra health care costs the employer is paying to the decreased productivity of the overweight employee. The issues of being obese and overweight have been linked to stress in the workplace and developing bad habits of eating and lack of physical fitness. A problem that has been prolonging obesity is the lack of availability of healthy foods at work (Berry et al., 2006). Healthy choices are not being made in the workplace. Cafeterias provide an abundance of junk food. Vending machines are full of soda pop and candy. It is hard to overcome the desires that the body craves when this type of food is present. Through the years we have learned bad eating habits and our society has created a fast food-eating nation. It should be a corporate priority to help these workers improve their well-being through health promotion. Such actions by the company will in turn improve more than the performance of the business. “A worksite health promotion program that is feasible, enjoyable, and effective may have significant benefits not only for employees, but for employers and society as well” (Racette et al., 2009, p. 108).

As a nation, we have created an epidemic of obese and overweight workers. It has been proven to be cost effective for businesses to implement health programs within the organization, and it will also lead to less turnover and more loyal employees (Anderson et al., 2009, p. 353). The workplace has the best atmosphere to influence employees to make the right decisions about good health and preventing diseases that cause many health risks.

“Obese adults incur annual medical expenditures that are 36% to 37.4% higher than those of normal-weight adults...The increase in medical expenditure costs is from more office visits for comorbid conditions, inpatient and outpatient hospital care, and prescription drugs” (Boardley and Pobocik, 2009, p. 243). All of these costs add up quickly. This is why a workplace wellness program is so vital to a company that wants to improve its workforce and bottom line. Just the increase in medical expenditures with obese employees should be a marker for the need for a change in health promotion within a company. “The number of lost work days is 13 times higher, and the cost of medical claims is 7 times higher among the heaviest employees compared to those of employees of normal weight” (Boardley and Pobocik, 2009, p. 245). Absenteeism is also very costly to the company. When employees don’t show up for work, production is slowed or even delayed. Workers need to be at their best when they are at work, and if they aren’t even there then the work doesn’t get done. The effects of these factors are costly and create a huge risk to the company.

Obesity affects a person’s life in many ways including both the quality and the length of life. Depending on the person’s race and sex, an obese person can have an 8%–22% reduction in length of life (www.cdc.gov). “Overweight and obese adults are also more likely to sustain injuries that require medical treatment. The odds of sustaining a medically treated injury are 15% higher for overweight adults and 48% for extremely obese adults. Medical conditions ranging from heart disease, diabetes, and cancer to now replacement surgery, depression, fatigue, and insomnia are all more common in obese patients” (Boardley and Pobocik, 2009, p. 247). According to these figures, obesity is hard to live with. Just knowing that a part of your life will be taken away because of disease is a hard thing to really comprehend.

Obesity can cause injuries in the workplace in which case the liability of the company is at stake. Workers’ compensation is a huge cost to many companies. If someone gets hurt on the job, then the company has to pay for it. Prevention and treatment of obesity is extremely important to the company and its employees. There are many factors that the company can control, but the health of their employees is something that has the potential to harm the company as a whole. There

are programs specifically designed to “not only improve health and wellness of employees, but also to identify employees at risk for work-related injuries, and then put them through a structured program with a personal trainer thereby decreasing their risk for injury...This program has demonstrated an ability to decrease overall worker’s compensation costs by improving an individuals’ capacity to meet job demands” (Hansen, 2010).

As company costs go up, revenue is taken away. Companies are going to be greatly affected by the increases in health care and also by the workers who are obese. Average annual medical expenditures are \$732 higher for obese than normal-weight individuals (Finkelstein et al., 2003). These are huge costs that employers are spending on the higher rates for their obese employees. They add up quickly and take away from the bottom line of the company. These dollars would be better used in preventative programs that help employees get to a healthy weight and lifestyle, rather than having the money go to medical costs that will just rise every year.

A company can also minimize turnover costs by implementing a well-being program for its employees. “Turnover costs can easily run between 1.5 and 2.5 times the annual salary of the incumbent job holder” (Wright, 2010, p. 17). That means that if a company has an employee making \$100,000 per year who quits, then the loss to the company is \$150,000–\$250,000. These costs come from the marketing of the job opening, application process, interviewing and hiring, new employee training, and many more processes. That is a very large expense for the company, and it would be worth it to the company to find ways to increase the well-being of each employee and retain the knowledge base that is already employed. People who are happy with their jobs are more committed to the company. Committed employees want the company to grow so they will perform to the best of their abilities. Happy employees equal happy customers. That is why companies need to make the health and wellness of their employees a top priority.

It is important to a company to improve and sustain the health of its employees. Employees know that a company is interested in their well-being when it provides support and opportunities to become healthier. The employee is an asset to the company and needs to be maintained physically and mentally. If people work hard and are not allowed a outlet to relieve their anxiety and stress, they will start having physical illnesses and mental distress.

Recreation

Recreational leisure is a proven way to alleviate the stresses and anxiety of work. Recreation activities like hiking, rock climbing, kay-

aking, and sports provide many short-term and long-term benefits to the participant. "Recreation activities are behavioral pursuits that are instrumental to attaining certain psychological and physical goals... People pursue engagement in recreation when a problem state exists; when an existing state does not match a preferred state" (Manfredo and Driver, 1996, p. 188). For instance, a person who is stressed by work and daily responsibilities may go hiking to attain an escape from the stress; however, there needs to be a further motivational factor for someone to do something like exercise or go for a walk outside. They need to have a beneficial view of the activity. "To the extent that a person has the required opportunities and resources, and intends to perform the behavior, he or she should succeed in doing so" (Ajzen and Driver, 1992, p. 207). For example, if a company provides passes to the local gym and creates an incentive for employees to use the facility, then the employees will have more intention of exercising and using the recreational facility. The short-term individual benefits of recreation (e.g., increased energy) lead to long-term individual benefits (e.g., increased self-esteem) and progresses to long-term societal benefits (e.g., quality relationships with others). The whole of society is improved by the small individual benefits that recreation provides.

Activities can be selected for the appropriate population in a company. Companies that want an effective wellness program need to set criteria with activities that motivate employees to engage themselves. Participants need to have an experience that has a positive outcome with enduring personal and social benefits. The program will not be successful if it does not include activities with these outcomes.

"Well-designed programs have the potential to extend beyond the worksite and positively influence dependents (spouses and children), and thereby reduce an organization's health care costs" (www.cdc.gov). Not only do wellness programs affect the individual but they affect the whole family. Recreational activities create benefits like self-confidence and build cognitive skills, environmental awareness, and many more that contribute to the welfare of the community. In this way, the benefits of the wellness program trickle down to society.

The children of today are going to be leaders of society tomorrow. Most children follow in the footsteps of the parents. If we want to change a social problem like obesity, we need to start with the parents. A wellness program will teach a parent the right skills to live a healthy life. That parent in turn will teach the rest of the family the skills and techniques learned from the programs at work. Everyone will be affected in a positive way through a well-designed workplace wellness program. It has the potential to affect the whole nation if people apply the principles taught and use them in their daily lives. An example

could be a decline in the obese population among children, adolescents, and adults.

Changing Behaviors, Physical Activity, and Eating Right

Since employers can benefit employees while helping the bottom line, they ought to do so. There is a need for businesses to become more involved in creating an atmosphere in the workplace that promotes physical activity and healthy eating. Companies can sponsor or create programs for their employees to enjoy while at work. An employer can offer healthy alternatives that will be appealing to employees. By implementing certain activities or products, organizations can construct a corporate culture that will change the behavior of their employees. As a domino effect, creating a healthy corporate culture will improve the lives of employees and increase company revenues.

According to Rippe and Hesse (1998), there are two epidemics facing the U.S.: the epidemic of obesity and the epidemic of inactivity. There are many benefits of physical activity that would be major factors in significantly decreasing both of these epidemics. Physical activity would be able to help restore the health and ability of those affected. There needs to be a change. One part of that change comes through the promotion of physical activity programs (Rippe and Hesse, 1998).

There are immediate benefits to having programs that help people learn how to be healthy and live a well balanced life. Employees will start feeling good about getting exercise and increasing their well-being. They will start realizing the immediate benefits of their participation and hopefully continue progressing through the program. These short-term benefits will, if the behavioral changes are continued, lead to long-term benefits. The long-term benefits act as a motivator for the employee to lead a healthier life. As behavioral changes are continued, the company may start seeing a difference in the way the employee works, acts, and feels. There could be a complete change in the people on health programs because there are multiple improvements that can take place. Individual benefits may include self-assurance and confidence, balanced living, stress management, sense of wellness, and a sense of control over one's life. These are important habits that employees need to have and maintain. The way one eats is also a habit that can be changed.

Larger companies can also have an impact on what kinds of foods their employees eat if the company offers a cafeteria for workers. Some companies have done interventions where they replaced unhealthy items on the menu with "the introduction of new low-energy-density

(ED) foods and provision of labels for all foods sold at lunch, which listed ED, calories, and macronutrient content” (Lowe et al., 2010, p. 144). With the introduction of labels that list important nutrition information, employees are able to keep track of the amounts of calories they eat and track their progress. The cafeteria intervention can also subsidize price of the healthy food, making it more economical and giving an incentive for the employees to eat healthy food. Giving employees the opportunity to make healthy food choices is vital because they will have the access to the important foods and will be able to know which foods are good. A cafeteria intervention paired with dietary and nutrition classes also have impact on the success of a health program.

Nutrition

Hippocrates stated “Let thy food be thy medicine and medicine be thy food” (Biziulevicius and Kazlauskaite, 2007, p. 712). The foods we eat affect our bodies. If we are to beat obesity and the rapid increase in diabetes, we need to change the way we eat as a nation. There needs to be a shift in our paradigm. Our food system is built to make money, not provide the nutrition to keep us healthy. Why are we as a nation so obsessed about cheap food? We are what we eat. So if we buy cheap, fatty food, what does that make us? There should be healthy, nutrient-rich food available to the public so we can stay healthy and not have a problem with cancer, heart disease, and cardiovascular diseases. Our nutrition is an important part of increasing healthy living in America. When we eat the right foods with vitamins and minerals, they enable the body to heal. We need to invest in the foods that will be the best for our bodies.

It is hard for healthy foods to compete with the price of unhealthy foods. Americans are obsessed with getting a good deal, but we don’t think about the after-effects of what we eat. If a person can eat cheaply at a fast food restaurant for the price of \$3–\$5, many will do that even though such a diet is not healthy. We are worth more than just filling our bodies full of junk foods that do not contain any nutrients. Our bodies deserve better. We should be investing in ourselves, helping ourselves be healthier and fit. There should be a point where enough is enough. Our bodies are worth more than the junk food we put in them. High-quality food may not be affordable to many. That may be one of the causes to the obesity epidemic. People are unable to feed their families the foods they need to be healthy.

A company can encourage employees to eat right by providing healthy options in the workplace. Larger companies can provide

healthy choices at the company cafeteria. Smaller companies don't have the luxury of having a cafeteria on the premises so they need to use simpler ways of helping employees have healthy food options in the workplace. A healthy refrigerator would give employees access to healthy snacks that they can purchase at a discount. It is a good alternative to the expensive and unhealthy snacks in the vending machines. If an employee forgets to bring lunch or is hungry during the work day, they know where to get a healthy snack that will be good for them.

Corporate Culture

Each company or business has a culture that portrays its values. The corporate culture can adopt healthy practices, such as physical activity, and incorporate them into the lives of its employees. To increase physical activity a company could encourage the use of onsite facilities, offer employee exercise breaks in place of the famous "coffee break," have lunch exercise classes, designate walking trails or make stairwells safe and attractive for the use of exercise. It is vital to motivate an employee to make healthy lifestyle changes. One way is to use some sort of incentive to get them going and staying on the right track. Financial incentives are known to be the best ways to reward effective behavioral change (Frølich et al., 2007). Money gets peoples' attention. If there is a financial incentive to workout and make lifestyle changes, people will do it for a cash bonus. Employers are starting to see the motivating factor and are implementing programs that have financial rewards.

For an employee, a financial incentive may be the main benefit at the start of a health promotion program. As they continue with the program, they may start to realize the importance of living a healthy lifestyle and will need to change certain habits. Their mentality of why they participate may change to being the healthiest for the benefit of their family or self and not just for the money. The financial incentive is there to get them on the right path. Once they get on the path of changing their behavior, it will be more motivation for them to live healthy because they choose to rather than just for the cash bonus.

The financial incentive for the employee can be large or small. That is to be determined by the company. For health promotion programs to succeed, they need to show that they have an impact on cutting cost and are a smart financial incentive for the company to make. "The ability of interventions to generate meaningful effect sizes is paramount to documenting health-related outcomes and economic (medical and productivity) impact" (Pronk and Kottke, 2009, p. 316). The documentation of every aspect is important to make sure that the programs are profitable in many different areas.

Many companies have realized that health intervention programs have been beneficial in many different ways, more than just by decreasing health care costs. Health programs need to demonstrate through research that there are many benefits to the company or they will not stay. If it is not cost effective to keep a program running, then there is no chance for the program to keep running. The owners of companies and the stockholders need to see what wellness programs provide and that they will benefit in the long run. They need to see financially how health programs cut costs. In the short term, the costs of implementing a health and wellness program may be more than was expected, but in the long run, these costs will be turned into gains and increased revenue (Burke and Logsdon, 1996).

Costs to the Company

Health care is a costly but important benefit many companies provide for employees. “Studies consistently show a positive correlation between employee obesity and increased health-care costs and illness related absenteeism in organizations” (Berry et al., 2006, p. 343). These health-care costs add up really quick when a third of the company’s employees are overweight and need more medical attention. In 2006, General Motors Corp. estimated that its obese workers “incurred an additional \$1,500 a year in health-care costs compared with employees of normal weight” (Berry et al., 2006, p. 343). With companies already paying increasingly large amounts of money on health insurance for their employees, they probably don’t like having to pay more for their overweight workers every year. The rise in obese workers is a concern for employers.

There is a need for more physical activity in the workplace. This is why employees need access to recreational facilities and resources that will help them achieve a balanced life. Employees who work in a sedentary environment the whole work day and have no access to fitness facilities or programs will have a greater chance of becoming overweight. “The true cumulative costs of obesity are incalculable because they are incurred in so many forms, including physical illness and mental distress, lower productivity, and increased healthcare spending” (Berry et al., 2006, p. 352). Health programs may cost money to run, but they can reduce the amount spent on overweight employees. The employee’s health will improve, and the cumulative costs of unhealthy lifestyles will decrease. It seems like a win-win situation. We will see through the following examples of companies how some health programs have worked and how some may need improvement.

Statistical Analysis Systems (SAS)

The analytical software company SAS believes that employees need to be healthy, happy, and productive. The owner of SAS, Jim Goodnight, believes that if employees are well motivated and happy and work in a creative environment, they tend to produce great products and make sure they produce the best quality in their work (www.sas.com). By having employees that strive to do the best that they can, the customers of the company are very pleased with their services and come back very satisfied. When you have happy customers, this makes the company happy because customers continue to buy the products.

The company has a 35-hour work week that has been the same since the company started. On the company campus, there is a medical facility that provides medical and financial benefits to both the employee and the employer. The employee saves time by going to the medical facility on campus rather than having to clock out and taking a couple of hours to get checked out and then driving back. With the facility so close and free of charge to employees and their families, they use these services more often, which helps create a healthier employee. It provides a better lifestyle through regular check-ups and ease of using the facility. Having the benefit of an on-campus medical facility is very valuable to the company by making health services available for their employees. This keeps the employee at work without missing too much time.

The overall work life is a concern for the owner of SAS. Mr. Goodnight has created “an environment that encourages creativity and innovation while allowing employees to balance work and life” (www.sas.com). He wants his employees to have a good balance between family and work so they can be productive in the office and successful at home. Some say that the reason that public companies do not provide an abundance of benefits like SAS is the pressure from Wall Street to please shareholders by delivering an increase in quarterly earnings. Most company boards are concerned about the cost vs. benefit ratio. They are more focused on increasing revenue than increasing the health of company employees. Fringe benefits are shrinking at most public corporations, but at SAS they care about their employees and set a higher standard for other companies by providing a wide selection of benefits.

The company does not have a set maximum amount of time for sick leave. When employees are sick, the company wants them to get better as soon as possible. They trust their employees to do what is right for the company, so if they are not well then the company wants

them to get well and take care of themselves. Also, if family members are sick, SAS allows certain employees to work from home online if it is applicable to their line of work. This allows the parent to have the flexibility to care for family and also work from home when needed. Using SAS's medical facilities for family members is also easier for the employees. They know that it is close and free of charge for general use. There are many facilities on the company campus for the use of employees.

At SAS, there is a 66,000-square-foot, state-of-the-art fitness center where employees can metabolize fat, pump iron, shoot hoops, and exercise even during working hours. There are an endless number of opportunities available for the employee to participate in recreational activities for their health. They are all available for free to employees and family. The company wants its people to enjoy their work so they are not stressed or in pain. With the availability of these facilities on the SAS campus, employees are able to enjoy them without having to drive somewhere to workout.

Employees who are parents like the on-site daycare that is provided at a third of the cost of normal daycares. At lunch time, parents can go and bring their kids to the company cafeteria where they can enjoy their meal together. This eases the stress of the parents knowing that their children are so close and they can spend time with them even at work.

Providing healthy food for employees at a discounted price is an added benefit at SAS. This is an incentive for them to purchase the healthier options. Giving employees healthy options enables the employee to make a healthy choice. Coinciding with wellness programs and nutrition classes, discounted foods at the cafeteria allow employees to follow dietary instructions they learn in their classes. This is a good way to motivate employees to be actively involved in their dietary habits. Empowering employees to make their own decisions can increase their fulfillment in their work experience.

If employees are satisfied with their job and their work environment, they will be loyal to a company that treats them well. This is why SAS has an employee turnover of less than 3% in an industry that averages 26% turnover (www.sas.com). By reducing this turnover, the company is able to retain millions of dollars. Employees are satisfied with their jobs and are productive during their seven-hour work days. They have access to an onsite health facility that is free of charge. There are wellness programs that they can participate in and can take instructor taught fitness classes at the facility. The well-being of the employee is very valued at SAS and it shows through the low turnover rate. Employees are very loyal to the company and never want to leave.

By creating a good work environment SAS is able to have high knowledge retention and cut on the costs of training and recruiting new employees.

Employees at SAS want the company to succeed so they do their best and perform at a high level. Employees know that they are privileged to work there and are grateful for the opportunities they have. Through the example of this software solutions company, having an employee-focused business plan will increase overall productivity by creating different faucets of healthy living. They have been able to construct an incentives program to reward employee wellness. The culture of the company keeps employees content and keeps them from leaving. If someone had the chance to work for SAS, they would not want to leave because no other company provides quality benefits like SAS does. SAS is the number one company to work for, ranked by CNN's money.com.

UVU Wellness

Here is a look into the wellness programs available here at Utah Valley University (UVU). At UVU Wellness Service, they help students get on track to living a healthy life style. The lifestyle of a student in college can be very abnormal with consuming junk foods and staying up late into the night. Many students live on a small budget, eating unhealthy foods, and fail to exercise because of the demands of work, school, and homework. For many students, controlling their weight is a problem. There is a popular saying that freshman gain 15 pounds their first year in college. This is because of poor eating habits, lack of vigorous exercise, stress, lack of sleep, etc. Freshmen are usually fresh out of high school, and this is their first time away from home. Many students do not know how to cook or even know which foods are good for them. Many buy what is cheap or what is fast and easy. Most students who go to college need direction and guidance with aspects of their lifestyle. UVU wellness services are available to students who are enrolled in classes, to all faculty members, and to staff.

At the beginning of the program, a participant is evaluated by taking a health risk appraisal. "The Health Risk Appraisal (HRA) is a measurement of one's current health status. It helps individuals become aware of the health needs and lifestyle practices that determine personal well being. Positive reinforcement of good health practices, along with recommendations for change (when needed), are made for each of the major health areas that are tested. Participants can complete a health risk appraisal by filling out an online lifestyle questionnaire, as well as completing 8 fitness tests" (www.uvu.edu). Through the test results, a

professional will be able to tell the participant what fitness and health level they are at currently. The professional has a better understanding of the results and knows what the participant needs to do to reach optimum health and create a sustainable lifestyle. It is important to point out behaviors participants do well and also the aspects they need to improve.

If an individual wants help or coaching as they go through lifestyle changes, they can participate in a program called Health Quest. It is “a personalized healthy lifestyle behavior change program. Participants will have the opportunity to meet with a wellness coach on a weekly basis. All materials used for this program are from accredited resources.” (UVU Wellness Website: <http://www.uvu.edu/wellnessed/>) It is a six-week program where they meet with a wellness coach every week, and they will have counseling sessions where they learn about the different aspects of changing their lifestyle behaviors. Having a coach to assist a person in changing bad habits is vital to the success of a program. It gives the participants responsibility to report to someone, and the coach is there to motivate them to succeed. The program covers the basics of nutrition and exercise and helps an individual know more about health behaviors. It is for individuals wanting a good start by learning about their health and how to improve it.

The health program at UVU concentrates on weight management, physical fitness, nutrition, controlling blood pressure and cholesterol, and managing stress and sleep. These areas of concentration are crucial to the lifestyle of a healthy person. For a college student each of these areas is important to maintaining a healthy lifestyle and also to doing well in their education. School and work take up most of the time, so having to concentrate on eating right, exercising, not stressing out over school and relationships, and getting the appropriate amount sleep is hard. Bad habits can be created when students attend college. Giving them the direction they need is extremely important. Not very many students take advantage of these resources because they either don't know about them or they feel these opportunities are not important to them or their future.

Wellness programs like the one at UVU need more publicity and marketing. Prospective participants need awareness that the programs are available to them. It is hard for a person to come to a meeting or take an evaluation if they do not know about it. In the case of UVU, there should be more on-campus information boards and emails about certain things individuals can do to improve their health and where to go to receive information. The available information should be easy to access. Awareness to future participants is crucial to the progress and success of a wellness program.

The UVU wellness program also has professionals come on campus and speak, presenting on a variety of health- and wellness-related topics. There are also workshops available to demonstrate the importance of creating a healthy lifestyle. There are many resources available to students and staff at UVU. People need to take advantage of the resources that are there for them to use. Each and every one of us needs to develop healthy habits that will create a sustainable and enjoyable life.

Public Employees Health Program (PEHP)

The State of Utah maintains a health insurance system for public employees. PEHP is a non-profit trust that ensures that public employees are given the choices of having health insurance available at an affordable price. This is a company that gives many options to government workers and provides them with many benefits. Their services are only available to the public sector (www.pehp.org).

There is an added benefit to being with this program. There is a free wellness and value-added benefits package that allows users to access a wide variety of lifestyle assistance. The program is called *pehpPLUS*. The goal is to make a healthy lifestyle more affordable. As a member, a person is “entitled to savings and discounts on an assortment of healthy lifestyle products and services” (www.pehp.org). Members are given discounts on health clubs, massage therapy, vitamins and supplements, weight management, life coaching, chiropractic, and many more. There is a wellness works program that “teaches ideas about managing your energy which is the key to healthy weight management. It also has dozens of fitness downloads, relaxation downloads, recorded advice from award-winning nutritionists and a full range of reporting so you can track your progress” (www.pehp.org). This is a great resource for everyone that needs help with any aspect of getting in shape. This information is vital to sustaining a healthy lifestyle and is available to those who want to use it.

There is an incentives program “that offers cash rewards for good health and healthy improvement” (www.pehp.org). Employees and members of their family can go to a primary care physician and get checked out. By filling out an assessment and going to the doctor, each member receives \$50 for taking the step to a better lifestyle. People can receive additional money for maintaining good health and those who are overweight or obese can receive money for lowering their body mass index (BMI). Cash rewards range from \$50 to \$300 for each step taken, depending on what the employee improves. These incentives motivate employees to take the necessary steps to change their lifestyle.

If these monetary incentives were not offered, many people would not use the resources available to them.

Offering these types of programs to employees puts the responsibility back in the hands of the employees. They don't have to use the programs, but they are there to use at the person's discretion. The company is not forcing the employee to get healthy but is assisting them if they need help. It makes the employee accountable for the actions that they take. It is vital for companies to make the necessary resources available so employees can make the decision to use them. If an employee is forced to do something, then there will be consequences in the long run. The individual needs to make the choice to participate or not. That is why the incentives are there—to motivate employees and give them a reason to use the services.

Rising Cost of Health Care

Good health is good for business. It keeps workers healthy and more productive, reduces absenteeism, improves employee morale, and enables the employer to recruit the best and brightest work force. According to the U.S. Census Bureau, 62% of Americans are insured by their employers. In 2002, the American Journal for Health Promotion stated that there is a 3.5:1 savings-to-cost ratio for work site health promotion programs. According to the Art of Health Promotion, in 2003, there was a 6 to 1 return on investment when combining reduced cost and improved productivity. It is important to have a sustainable workforce that is healthy and operational.

All companies want to improve their bottom line. Prevention appears to be more promising than continuing to pay medical costs for the diseases that are preventable, especially those diseases that are caused by poor nutrition and lack of exercise. The medical costs that companies are paying for are impairing their competitiveness in the market. Companies are paying the costs of obesity, either directly by paying the medical costs or indirectly because of the reduced productivity of the obese employee. It is essential that employers invest in prevention because they will not be able to afford the rise in medical costs.

Incentives

Why should your company have a wellness program? You may be thinking that you don't need the extra expense, the hassle, the paperwork, especially in this economy. You may be a small company that is too busy to get involved in a wellness program. Would a wellness program even be feasible for your company? Programs come in many sizes and have been clearly shown to increase productivity, improve em-

ployee morale, lower health insurance costs, lower absenteeism rates, and provide a number of other advantages.

Bad habits create an unhealthy lifestyle, and it is hard to break old habits. Even if there are resources available to an obese person, they most likely will not take advantage of them because of the habits they don't want to change. There needs to be more incentive for the obese person to change and the normal person to maintain a healthy lifestyle. The obese person struggles with the knowledge that they are going to die at an early age if they don't change, but apparently Americans don't care. They are still becoming obese. Americans are conscientiously participating in a culture that is increasingly getting fatter. How can organizations and companies change the way America lives? They can do it by using the number one motivating factor for people. Money is the one thing that grabs the attention of everyone. That is why there needs to be a monetary incentive for most people to go and do something about their health. This is how companies will be able to motivate their workers to get healthy and be more productive. If there is no incentive, then there will be no change. With incentives, the change will occur faster and be more effective.

Return on Investment

It is important for a company to show that its wellness program works and that there is a return on the investment. If the program doesn't show that it supports the objectives of the company and increases revenue in the company, then the plan will not work. Human Resources needs to develop wellness programs that have a positive cost-benefit analyses to show that there are many benefits to the programs beyond just having revenues increasing. The benefits need to outweigh the costs of the program.

Employees will feel appreciated by the company providing a way to design benefits for them to increase their well-being and improving their own lives. Companies need to encourage the employees and reward them for engaging in these wellness programs. The programs should provide the employees with opportunities and resources that will help enable them to progress and complete objectives. A rewards system is vital to the implementation and completion of these physical activity interventions. Employees need to see the benefits that they will receive for the participation and completion of these programs.

The return on investment won't only be seen by the company but also by the individuals participating in the programs. The company will be increasing its revenue from a higher-producing workforce, less absenteeism from illnesses and injuries, healthier and happier employees

with fewer turnovers, and spending less money on health care costs. The company's human capital is very important and should be taken care of and handled well. The employees are an asset to the company and there should be an investment in their health and well-being. It is important to have employees operating at their optimal ability and that is accomplished by having them go through these physical activity and wellness programs. It is a way to improve the lifestyle of the employees in and out of the workplace.

The benefits of the programs influence the employees not just at work but also in their personal lives. The benefits tend to offset the direct costs of medical care. "Indirect costs are often far more important than the direct medical care costs" (Popkin et al., 2006, p. 271). Saving money on health care costs is not the sole purpose of these health promotion programs. It is hard to measure the increase in self-confidence and satisfaction with their jobs. There are many indirect benefits the employees receive by putting effort and time into the programs. Incentives may offer another alternative to motivating the employees to achieve certain goals. When they reach those goals, the company can reward the employees with recognition and give them a monetary incentive.

Recommendations

Companies that are in the process of implementing a wellness program need to create goals and objectives they would like to achieve. Our recommendation for them is to start their programs step by step. It would be most beneficial to survey the employees and get opinions about what they would like to have in a wellness program that will be beneficial to the employee. It is important to get the input of the employee because they are the ones who will be using the program. If they are not interested in the activities, they will not be motivated to participate in them. Implementation of wellness activities that meet employee needs and interests is important to the success of a program.

Setting company, department, and individual goals with a reward system will get people motivated to reach their goals. With goals in place a company can change the environment of the workplace in a way that supports a healthy lifestyle. Changes can include providing healthy choices for lunch or having a fitness area onsite. A company can even have an "employee of the month" who is featured for accomplishing certain program goals. Acknowledging the efforts of employees can improve employee morale to work hard and reach their goals.

To increase physical fitness, a company can partner with a local fitness club and subsidize the cost of a membership. Employees would

have access to a gym and personal training, if needed. Having support from the company and from coworkers, employees can have a support network that will help them accomplish their goals. Having someone to workout with can be a motivating factor for the employees. By using some of the ideas presented in this paper, a company can provide a successful and beneficial wellness programs to their employees.

Conclusion

When employees know that their employer cares about their life and how they are affected by their obesity, the employees will have a better understanding and appreciation for their company. Employees stay with companies that care about them and treat them like real people with feelings and needs. People want to be cared about and recognized for their efforts. A company can show its recognition of employees and give them praise for the good job that they have done.

We could keep going down the same path of letting employees become overweight, which will lead to many types of disease and costs to the company. Companies do not have to implement wellness programs into their corporate policies. It is not required of any company to do so. If we want to have a nation that is overweight and draining the company and the economy, then we can just let things proceed without any intervention. The rise in workplace obesity will continue if there is no change in corporate policy and culture.

There needs to be a change. Corporations need to take the initiative. They are in the best position to make the largest impact on the greatest amount of people. By changing the behaviors of employees, there will be a chain effect that will be felt not just by the company but by the whole nation. Families will be influenced by the examples set by the heads of households. We will see the behavior in children start to change. By helping out the company, by cutting costs, by implementing health and wellness programs, and by creating a corporate culture that upholds physical activity and healthy eating, we will see the benefits trickle down throughout society.

Resources

Ajzen, I., Driver, B. L., (1992). Application of the theory of planned behavior to leisure choice. *Journal of Leisure Research*, 24(3), 207.

Anderson, L.M., Quinn, T.A., Glanz, K., Ramirez, G., Kahwati, L.C., Johnson, D.B., et al., (2009) The effectiveness of worksite nutrition and physical activity interventions for controlling employee overweight and

obesity. *American Journal of Preventive Medicine*, 37 (4), pp. 340-357.

Avena, N.M., Rada, P., Moise, N., Hoebel, B.G., (2006) Sucrose sham feeding on a binge schedule releases accumbens dopamine repeatedly and eliminates the acetylcholine satiety response. *Neuroscience*, 139 (3), p. 813-820

Berry, L.L., Seiders, K., Hergenroeder, A.C., (2006) Regaining the health of a nation: What business can do about obesity. *Organizational Dynamics*, 35 (4), pp. 341-356.

Bianchini, F., Kaaks, R., Vainio, H., (2002) Overweight, obesity, and cancer risk. *Lancet Oncology*, 3 (9), pp. 565-574.

Biziulevicius, G., Kazlauskaite, J., (2007) Following Hippocrates' advice 'Let food be thy medicine and medicine be thy food': An alternative method for evaluation of the immunostimulatory potential of food proteins. *Medical Hypotheses*, 68 (3), pp. 712-713.

Boardley, D., Pobocik, R. S., (2009) Obesity on the rise. *Primary Care: Clinics in Office Practice*, 36 (2), pp. 243-255.

Burke, L., Logsdon, J. M., (1996) How corporate social responsibility pays off. *Long Range Planning*, 29 (4), pp. 495-502.

Finkelstein, E.A., Fiebelkorn, I.C., Wang, G., (2003) National medical spending attributable to overweight and obesity: how much, and who's paying? *Health Affairs*, W3-219-226.

Frølich, A., Talavera, J. A., Broadhead, P., Dudley, R. A., (2007) A behavioral model of clinician responses to incentives to improve quality. *Health Policy*, 80 (1), pp. 179-193.

Hristova, M., Aloe, L., (2006) Metabolic syndrome – Neurotrophic hypothesis. *Medical Hypotheses*, 66 (3), p. 545-549

Hansen, K., Email to the author, November 3, 2010.

Lowe, M.R., Tappe, K.A., Butryn, M.L., (2010), An intervention study targeting energy and nutrient intake in worksite cafeterias. *Eating Behaviors*, 11 (3), p. 144-151

Manfredo, M., Driver, B. (1996). Measuring leisure motivation: A

meta-analysis of the recreation experience preference scales. *Journal of Leisure Research*, 28(3), 188.

Popkin, B.M., Kim, S., Rusev, E.R., Du, S., Zizza, C., (2006) Measuring the full economic costs of diet, physical activity and obesity-related chronic diseases. *Obesity Reviews*, 7 (3), pp. 271-293.

Pronk, N.P., Kottke, T.E., (2009) Physical activity promotion as a strategic corporate priority to improve worker health and business performance. *Preventive Medicine*, 49 (4), pp. 316-321.

Racette, S.B., Deusinger, S.S., Inman, C.L., Burlis, T.L., Highstein, G.R., Buskirk, T.D., Steger-May, K., Peterson, L.R., (2009) Worksite opportunities for wellness (WOW): Effects on cardiovascular disease risk factors after 1 year. *Preventive Medicine*, 49 (2-3), pp.108-114.

Rippe J.M., Hess S., (1998) The role of physical activity in the prevention and management of obesity. *Journal of the American Dietetic Association*, 98 (10 Suppl. 2), pp. S31-S38.

Sturm, R., (2007) Increases in morbid obesity in the USA: 2000-2005. *Public Health*, 121 (7), pp. 492-496.

U.S. Department of Health and Human Services. The Surgeon General's Vision for a Healthy and Fit Nation. Rockville, MD: U.S. Department of Health and Human Services, Office of the Surgeon General, January 2010.

Wright, T.A., (2010) Much more than meets the eye: The role of psychological well-being in job performance, employee retention and cardiovascular health. *Organizational Dynamics*, 39 (1), pp. 13-23.

Yackobovitch-Gavan, M., Nagelberg, N., (2009) The influence of diet and/or exercise and parental compliance on health-related quality of life in obese children. *Nutrition Research*. 29 (6), pp.397-404.

The Challenge: Experiential Education in Theory and Practice

Peter Robinson, Laurent Josien, and Rachel McGovern
Utah Valley University

Abstract

Because of the complex nature of entrepreneurship and the wide range of knowledge, behaviors, and motivations involved in the entrepreneurial processes, experiential education has emerged as the foundation pedagogy for developing entrepreneurs. It has become critical to provide theory-backed models and principles as a foundation for the pedagogy. This paper explores the foundations and structure of attitude theories and goes on to provide a model for the application of attitude theory in experiential education within the context of entrepreneurship education. The application of theory is illustrated through an in-class entrepreneurship simulation referred to as “The Challenge.”

Introduction

I hear and I forget
 I see and I remember
 I do and I understand
 --Confucius

The concept of entrepreneurship has been around us for a long time, yet entrepreneurship education is only a recent addition to university curricula.

Indeed, the entrepreneurship concept has recently been a major focus in Higher Education Institutions (HEIs) all over the world. In the U.S. and U.K., entrepreneurship classes are increasingly offered to students as part of their undergraduate and graduate choice in curriculum (Jack and Anderson, 1999; Morris et al., 2001; Klappa, 2004). Entrepreneurship majors have been available in U.S. HEIs since the early 1980s (Hills, 1988). Morris et al. (2001) found that the number of business schools in the U.S. providing classes in entrepreneurship or new venture creation has grown from roughly 25 in 1980 to over 700 by 2000. Bennett (2006) also reports a major growth in master's level provisions of entrepreneurship programs. Many universities now offer entire undergraduate, graduate, and even doctoral degrees in entrepreneurship or "business enterprise" (Adcroft et al., 2004); however, that increase in offering was not without some developmental issues. For instance, Sexton and Bowman (1984) complained that "the content of a typical entrepreneurship course varies according to the teacher's personal preferences as to definition and scope" (p. 21). Furthermore, Gorman et al.'s (1997) literature review of entrepreneurship education noted very little uniformity among the courses offered, which was confirmed by Shane and Venkataraman (2000), who found wide variation in course content. They advance that the reason for such variety is rooted in the absence of a clear consensus on the definition of entrepreneurship and the lack of generally accepted paradigms or cohesive theoretical frameworks in the entrepreneurship education area.

Nevertheless, some advances have been made in entrepreneurship education, and tools have been developed to enhance the teaching of entrepreneurship. Since the 1980s, two major schools of thought in entrepreneurship education have emerged: (1) a skills-based approach and (2) an attitude-based approach (Bennett, 2006).

The first approach is based on the skills set needed for entrepreneurs. Skills-based programs seek to teach students the mechanics of running their own business. They tend to be "highly structured, consensus-oriented and unstressful" (Sexton and Bowman, 1984, p. 21) and

usually involve instruction on how to raise finances, how to choose one's location, taxation, employment, legal requirements, entry level book-keeping, and other basic instructions.

Teaching methods usually include case studies, lectures, and assigned reading intended to develop the student's critical judgment and capacity to digest, understand, and analyze information (Collinson and Quinn, 2002; Davies et al., 2002; Ladzani and Van Vuuren, 2002). Assessment and coursework typically comprises written reports and the development of a business plan (Hills, 1988). Courses of this nature are said to be popular because enrollees frequently desire practical, highly specific, and "hands-on" information about small business management issues (Collinson and Quinn, 2002; Ladzani and Van Vuuren, 2002).

Hills' (1988) survey of entrepreneurship education programs in U.S. universities found that instruction in small business management processes was the primary activity of most courses. An important justification for a university deciding to run this kind of program is the substantial body of evidence that exists suggesting that new businesses rarely fail because their owners lack innovation, self confidence, imagination, etc., but mainly in consequence of their owners' ignorance of management, marketing, finance, budgetary control, employee recruitment, and other aspects of personnel administration (Hambrick and D'Aveni, 1988; Jansen and Van Wees, 1994; Davies et al., 2002, Ibrahim and Soufani, 2002; Peterman and Kennedy, 2003; Keogh and Gallaway, 2004). According to the skills-based approach, entrepreneurs are "born" entrepreneurs, and entrepreneurship education should help entrepreneurs in supporting skills like marketing and management.

The second approach described by Bennett (2006) is an attitude development approach. The attitude development approach emerged from criticism of the skills-based approach. Some scholars view the skills-based approach as "passive," "mechanistic," and in contrast "with the reality of the entrepreneur operating with intuition and limited information under acute time pressure" (Henderson and Robertson, 1999, p. 238). Rather than focusing on a "best-practice" pedagogy, critics suggest that entrepreneurial education should try to "inculcate the necessary attitudes, values, and psychological sets" of the successful entrepreneur (Curran and Stanworth, 1989, p. 13) and develop appropriate personal attributes such as innovativeness; the willingness to take risks, to fail and start afresh; creativity; and determination and self-direction (Garavan and O'Cinneide, 1994; Jansen and Van Wees, 1994; Hynes, 1996; Engelen, 2002; Gibb, 2002; Deamer and Earle, 2004).

As a result, and because of the complex nature of entrepreneurship and the wide range of knowledge, behaviors, and motivations in-

volved in the entrepreneurial processes, experiential education has emerged as the foundation pedagogy for developing entrepreneurs.

Table 1 summarizes the differences between the two approaches developed above.

Table 1. Characteristics of skills-based and attitude-based entrepreneurship education		
	Skills-Based Approach	Attitude-Based Approach
Philosophy	Entrepreneurs are born with critical characteristics	Individuals develop entrepreneurial characteristics through experience
Knowledge orientation	Business-based content (facts and concepts)	Business processes and content
Locus of instruction	Teacher centered	Student centered
Pedagogy	Case studies, lectures, and assigned reading	Experiential and engaged learning
Objective	Identify and analyze critical information, understand concepts	Identify and analyze critical information, apply concepts
Unit of analysis	Firm and firm performance	Individual and individual differences
Learning focus	Practical, highly specific information on business formation and management	Practical, hands-on experience on identifying, evaluating, and executing business opportunities

Because some would criticize experiential education as lacking rigor, it has become critical to provide theory-backed models and principles as a foundation for the pedagogy. As we explore and develop better theories and models, our application in education will become more rigorous and effective in helping students become more entrepreneurial: rigorous because we will understand why and how education and experience influence entrepreneurial propensities, and effective because we will understand how to structure activities to influence the desired outcomes.

Because of the complexity involved in entrepreneurial activities, developing entrepreneurs in an academic setting requires an equally complex set of activities to facilitate the acquisition and understanding of the requisite entrepreneurial abilities (Robinson and Malach 2004). Teaching entrepreneurship, then, requires a multi-dimensional and cross-disciplinary approach with an emphasis on dynamic processes that will expose students to the complexity of entrepreneurial activities in such a way that their actions can be examined and understood relative to the context of their own entrepreneurial development. Experiential education takes many forms (Gibb, 2002) including but not limited to:

- Case analysis and writing
- In-class and out of class exercises
- Simulations
- Projects
- Interviews
- Business plan writing
- Consulting/SBA cases/service learning

Experiential education provides a range of activities for teaching complex concepts and activities. This paper explores a theory base for the development of effective experiential educational for entrepreneurship; it then describes the implementation of that theory base through the use of “The Challenge,” an in-class exercise successfully implemented in entrepreneurship education across multiple institutions for over 20 years for various audiences ranging from high school students to MBA students.

The foundations for good pedagogical models are good theories and the foundations for good theories are the paradigmatic assumptions upon which they are built. If the foundational assumptions are inconsistent with the desired pedagogical outcome, the model will also be inconsistent and the results will fall short of expectations.

The intended outcome of experiential entrepreneurship education should move beyond students’ understanding of concepts, principles and practices to a shift in students’ mindset toward being more entrepreneurial. Mindset is defined as:

A set of attitudes held by an individual that influences or predisposes them to interact (perceive and behave, act and react) with the world in consistent ways. It is synonymous with intentions, inclinations, a frame of mind, an approach, or an outlook (Robinson, 2010). A mindset that is more entrepreneurial is one that would facilitate the identification, evaluation and execution of business opportunities (Shane and Venkataraman, 2000).

Foundational Assumptions

Robinson (1996) noted that social science theorizing about the nature of personal characteristics has focused on the dynamic interactive relationship between an individual and the environment. Mitchell and James (1989) describe this relationship as:

A new view that stresses the important attributes of people, their contexts, and their interactions. First, the person is seen as fitting into an environment. Second, both the person and the environment change over time. Third, changes in the person can cause changes in the environment and changes in the environment can cause changes in people. Fourth, people are both active and reactive with respect to these changes. Fifth, people's views of their past and future influence whether they are active or reactive and how much or how little they change. Thus, what emerges is a human who is active psychologically and behaviorally, interacting in a dynamic way with a changing environment. For the person, there is both stability and change, there is [pro]active and reactive behavior, [and] there are abilities and acquired skills that merge. (p. 147)

This view of a dynamic interaction between people and their environment has strong implications for education and training where the objective is to go beyond rote learning. With a dynamic paradigm, learning becomes a process whereby knowledge and understanding¹ are created through the transformation of experience with a realistic environment (Robinson and Malach 2004). In education this is best achieved through actual performance of tasks relevant to the learning objectives (Specht and Sandlin 1991).

The Theory Base

The theory base that forms the best foundation for the concept of mindset is the tri-partite attitude theory. Attitudes are a predisposition to respond in a generally favorable or unfavorable manner with respect to the “object” of the attitude (Ajzen 1982; Huefner 1991; Kristensen et al. 2001; Robinson et al. 1991; Rosenberg and Hovland 1960; Shaver 1987). Attitude is a dynamic construct, changing over time and acquired through experience. The “object” could be as concrete as a spe-

¹ Defined as a quality of knowing going beyond cognitions such as facts, concepts, beliefs and principles to include conation and affect.

cific individual or business or as abstract as a way of interacting in the business world. With regard to entrepreneurship, Robinson et al. (1991) demonstrated that a constellation of attitudes effectively differentiates between those who start businesses and those who do not. The embodiment of this framework is the Entrepreneurial Attitude Orientation developed by Robinson (1987). It is the most widely used and cited attitude measure in the field (Robinson, 2010).

A tri-partite attitude model holds that people interact with and learn from their environment in 3 distinct but related ways: affective, cognitive, and conative.² This model of attitude uses all 3 types of interaction in combination as 3 interrelated aspects of attitude (Allport 1935; Breckler 1983, 1984; Carlson 1985, 1994; Gibb 2002; Katz and Stotland 1959; Kothandapani 1971; Kristensen et al. 2001; Ostrom 1969; Robinson 1987; Rosenberg and Hovland 1960; Shaver 1987). Thus, attitudes are a combination of all 3 elements interacting with the perceived environment as well as with each other (Allport 1935; Bartal 1992; Breckler 1984; Gibb 2002; Kristensen et al. 2001; McGuire 1969; Smith 1947; Verplanken et al. 1998). The affective component consists of positive or negative feelings toward the object. The cognitive component consists of the beliefs and thoughts of an individual about an attitude object. The conative or behavioral component consists of intentions, predispositions, or inclination to behave in a given way toward the object. Thus, one can see the significant correspondence between the concept of mindset and the tri-partite model of attitudes.

Attitude Change and Education

Attitude theory and models of attitude change conceptualize the acquisition of understanding as a shift in the way an individual perceives and interacts with the world around them. This shift usually includes not only the cognitive aspects of the person–environment relationship but the conative and affective aspects as well. In short, attitude theory offers a framework for understanding the ways in which people are affected by experiential learning activities (Robinson 1996; Robinson and Malach 2004).

Attitudes, then, are the focus of one's affect, cognitions, and conations on real or conceptual objects in an individual's environment. The focus need not be balanced, with affect, cognitions, and conations having equal influence. Indeed, in traditional education with rote learning

² The aspect of mental life having to do with purposive behavior, including desiring, resolving, and striving (Webster's College Dictionary, 1991)

the emphasis is almost entirely cognitive. Based on models of experiential education, it can be assumed that the more balance between these three dimensions of our interaction with the environment, the more “reliable” the learning will be. Crosby (1995) indicated that the assumptions underlying experiential education are more reliable than those underlying traditional theories of education. By this she meant, “Students educated according to these assumptions are better prepared to deal with the world than are students educated according to traditional epistemologies” (pp. 4–5).

Applying this approach to education provides a foundation for experiential learning that not only encompasses the cognitive elements of traditional educational programs, but also includes affective and conative elements in the learning process. Learning can then be viewed as a process of attitude change or acquisition with cognitive, affective, and conative elements all influenced to some degree.

For example, education learning has consisted primarily of classroom experiences in which an instructor delivers a lecture or discusses a case. This process focuses primarily on the cognitive aspects of business, such as the acquisition of knowledge and beliefs about the subject at hand, with little thought given to the affective and conative aspects of the business experience. Although the cognitive and analytical aspects of business are important for success in most ventures, they lack emotional anxiety, excitement, and action imperative that emerges from the dynamic tension existing in the “practice” of business or more precisely entrepreneurship (Robinson and Malach 2004).

Viewed through the lens of attitude theory, education using the “practice” of business, focusing on a balance between affect, cognition, and conation should result in better retention, understanding, and application of the material being learned. Indeed, Specht and Sandlin (1991) and McMullan and Boberg (1991) have demonstrated that as pedagogy moves away from traditional teaching to more experience-based methods, the long-term retention of material, the understanding of concepts, and the application of knowledge are enhanced. Concepts take on a greater sense of reality, which creates a more relevant context for knowledge within the individual.

Experiential Education

When a task is performed it is referred to as experience. In experiential education, we often model the behavior of the actual task with the expectation that the cognitive and affective aspects of the experience will emerge along with the behavior. In other words, if we have students execute a simulation, they will experience the behaviors, cog-

nitions, and emotions emerging from the situation in a dynamic and reciprocal manner. At times their previous experience, along with attendant thoughts, beliefs, and feelings, will influence their behaviors in the new situation, and at times their behavior in the moment will elicit new thoughts, feelings, and behaviors.

The fundamental assumptions of experiential education apply to many types of activities. “Reliable” learning will emerge from the experience depending on the extent to which activities are consistent with the dynamic person–environment interaction critical to achieving a balance between affective, cognitive, and conative responses. Bringing experiential education into a concrete form requires a shift in the mindset of the individual, away from a passive receiver of knowledge to one where they are truly engaged with their environment. Joplin’s (1995) five-step model for how this engagement may be structured for maximum effect includes the following elements; chief among these is reflection in the feedback and debrief elements:

Joplin’s Five-Step Model

1. **Focus.** This consists of presenting the task and isolating the attention of the learner for concentration, preparing the individual for the challenging action that is to follow.
2. **Action.** This stage places the learner in a stressful situation where he is unable to avoid the problem presented, often in an unfamiliar environment requiring new skills or the use of new knowledge. Actions may be physical, mental, emotional, spiritual, or any combination of several actions. Actions are student centered, requiring a sustained effort in observing, sorting, ordering, analyzing, behaving, emoting, struggling, and, in general, interacting with the environment and the presented problem.
3. **Support.** Support provides security and protection for the learners, giving them the confidence to try new things and stretch themselves in risky situations.
4. **Feedback.** This provides students with information about the students’ performance on the task and gives some guidance for future actions.
5. **Debrief.** Here the learning is recognized, articulated, and evaluated by the learner through a process of reflection and self-examination. The teacher is responsible for seeing that the actions previously taken do not drift along unquestioned, unrealized, unintegrated, or unorganized.

(from Joplin 1995)

The Practice of Experiential Education

“Engaged learning” teaches complex principles through the use of structured behavioral activities where the learner engages with the environment in significant and meaningful ways (Dutton and Stumpf, 1991; Thatcher, 1990). These activities are significant in that actions of the learner have consequences in the immediate activity and potentially outside of the activity. The activity is meaningful when it provides students with a learning experience that is applicable to their long-term educational objectives. In general, the experiences are more significant and meaningful when they provide an opportunity affective, cognitive, and behavioral interaction with the environment. Robinson and Malach (2004) identified elements that enhance the effectiveness of attitude change and thus are elements of good engaged education.

Elements of Good Engaged Education

A balance of content and process. An examination of process is fundamental to experiential learning. How we solve the problem is as important as the solution in the learning process (Chapman, et al., 1995; Joplin, 1995).

Student rather than teacher based. The teacher will create safe working boundaries and then get out of the way, placing the responsibility for learning squarely on the student (Joplin, 1995).

Personal not impersonal in nature. The learner as a perceiving, thinking, feeling and behaving human being is placed under stress to perform in a specific situation. Individuals must be engaged in the process to bring all their faculties to bear on the problem. It must become personally relevant to them (Joplin, 1995).

Holistic rather than reductionist. Complex environmental situations facilitate learning the complexity of relationships in real world settings. Solutions must take into account the rich diversity present in realistic environmental situations. The whole is in reality greater than the sum of the parts (Chapman, et al., 1995; Joplin, 1995).

Creates an emotional investment. Any experiential learning model, that does not recognize the importance of emotional investment, diminishes its potential effectiveness for the learner. The processes need to engage the learner to a point where what is being experienced strikes a critical central chord within the learner (Chapman, et al., 1995; Joplin, 1995).

Reflection and self-examination. Experience without reflection and examination is only interaction and not education. Learners must reflect on the processes and examine their reactions to the situation and their actions (Chapman, et al., 1995).

Moving outside one's comfort zone. Experiential learning often stretches learners beyond their normal range of cognitive, conative, and affective responses. This enables them to examine their normal responses and learn new patterns of responding to situations (Chapman, et al., 1995).

(from Robinson and Malach, 2004)

In summary, engaged learning is an approach to education where tacit as well as explicit knowledge and understanding is gained through interaction with the environment. These experiences are complex because multiple thinking, feeling, and behaving beings are interacting in a complex environment. By examining these, relationships, new insights into the practice of any specific field can be developed.

In the field of entrepreneurship, there is a significant body of emerging research linking experiential education to positive changes in students' entrepreneurial mindset (Tam, 2009; Tam et al., 2009; Harris et al., 2007/2008; Rasheed and Rasheed, 2003; Souitaris, et al., 2007). This research demonstrates that experiential education, and entrepreneurship education, in particular, in addition to teaching concepts, principles, and practices, has a significant effect on the entrepreneurial mindset of the students. Tam's (2009) research demonstrated that:

Students became more entrepreneurial through the confidence gained from the real-world relevant content material and the application focus of the curriculum. The mix of having practice-oriented instructors capable of blending practice with theory, frequent exposure to reputable, practitioner guest-speakers with expertise in topical areas, learn by doing through hands-on projects in collaborative teams are contributory factors. (p. 3)

This paper will now explore a specific application of experiential entrepreneurship education in the form of an in-class exercise. This exercise is used as an example of how the principles of experiential education can be applied in a classroom setting and provides a basic outline for an exercise that may be used to teach complex principles that are difficult to teach through traditional methods.

The Challenge

Experiential learning activities are designed to teach complex principles through the use of structured behavioral activities (Dutton and Stumpf, 1991; Thatcher, 1990). The Challenge, as an experiential learning activity, involves structured scenarios within which the participants enact unstructured roles. Learning takes place through the examination of interactions between (a) the scenario and the roles adopted by the individuals, (b) the different roles adopted by the individuals in the activity, and (c) the roles adopted by the participants and their normal mode of behavior.

Engaged learning involves students in a situation with a compelling problem that must be solved; however, it is not the solution to the problem itself that creates the learning for the student but the processes involved in defining the problem, exploring multiple solutions, selecting the optimal solution, implementing the solutions, and finally resolving the problem. The examination of the process is the critical element.

In addition to the direct applications of concepts from the minefield in teaching entrepreneurship, there are several concepts that can be taught through the minefield that support an entrepreneurial attitude and understanding the complexity of enterprise formation. These concepts include: (1) team building elements, which are a success factor in any enterprise; (2) opportunity recognition aspect, which is a recognized characteristic of entrepreneurs; (3) business decision making in matching resource to the task and resource allocation; and (4) again the risk and potential rewards of opportunities and resource decisions.

A simulation is a simplified situation that contains enough similarities with reality to elicit real-world responses from the participants (Keys and Wolfe, 1990). Robinson (1996, 2010) explains that the Challenge, although it is not couched as an entrepreneurial event, can be related to the entrepreneurial process. First, the participants are part of a venture, an undertaking involving risk and uncertainty. In defining the scenario, Robinson (2010) goes on to introduce key variables consistent with an entrepreneurial situation, such as:

- A superordinate goal or mission to be accomplished
- Limited resources that may change based on factors within and external to the group
- An uncertain outcome including the possibility of failure
- An ambiguous situation where the procedures are not set but must be developed by the group
- Multiple options in terms of strategies and tactics in accomplishing the task

- Emotional involvement and commitment to reach the objective
- An action imperative where group members take action often without understanding the situation.

In short, The Challenge closely resembles a new venture along several dimensions that can be controlled by a facilitator. This is the essence of a business simulation. Other similarities between the exercise and new venture development may be drawn out in the debriefing session at the conclusion of the actual exercise.

The following description of The Challenge activity does not fully convey the affective and behavioral complexity of the activity. The only way to fully appreciate the execution and learning is to participate in the activity. Robinson (1996, 2010) provided the following description of The Challenge and the section on how The Challenge applies to engaged learning as a way of reinforcing and guiding practitioners who wish to use the exercise.

The Scenario

The members of the group should imagine that they have been caught in the Soviet Union and have arranged to escape from the “gulag” in which they have been held. They have hired or bribed a guard to help them get back to their own country. Their last obstacle is a very sophisticated electronic minefield. The minefield is constructed with sensors so that if one person touches any part of the ground within the field, that person and anyone else on the field or in the near vicinity is killed (failure).

To accomplish this seemingly impossible task, the guide has provided the group with a number of anti-mines disguised as rocks, blocks, or other flat unstable objects that are capable of sustaining 200 to 500 pounds of pressure, and can be easily picked up, carried, and tossed very short distances without breaking, chipping, or cracking. Rocks are often used as anti-mines, but if they are unavailable, scrap lumber of about one half square foot, cut at odd angles, may be used. Individuals can place the anti-mines on the field and stand on them without triggering the mines. Several people may share one anti-mine. However, if anyone falls or steps off the anti-mine, he or she is immediately blown up—along with the entire group—and will have to return to the starting line to begin again. The group should start with two or three fewer anti-mines than the total number of people participating in the exercise.

Three specific rules must be stressed. First, because the group is trying to get across the minefield without being detected, it must

do so in total silence; that is, no talking, laughing, or even whispering. Group members can communicate as long as it is not written or oral. Second, no one may cross over the finish line until everyone is on an anti-mine in the minefield; that is, out of the starting area. If one person crosses the finish line while another is still in the starting area, it completes a circuit that blows everyone up and the whole group starts over again. Finally, there are to be no sacrifices or martyrs; everyone must exit the minefield.

The Minefield

The minefield to which the group is taken is described to the participants as an area of unlimited width and variable length; in truth, it is usually an open area at least 50 yards long and 10 yards wide, ideally with grass (or carpet if indoors). At the onset, the field should be a minimum of 10 to 15 feet longer than the group could stretch if everyone in the group stood, arms spread, finger tip to finger tip, lengthwise on the field. The perception of the group should be that this is an impossible distance to cross given the resources available. The length of the field may change at any time during the activity, with the leader determining the length at any particular moment. A visible marker should be placed at both ends of the field to identify the beginning and end for the participants (see Figure 2).

Ideally, the activity should take place outside on a lawn or grassy area with access to rocks or other objects. It may also be conducted indoors in a large hallway or gymnasium. The leader should be creative in building environmental factors such as sidewalks, ditches, landscaping, buildings, hallways, and so on, into the scenario.

The Leader

The guide has the freedom to roam across the minefield without detonating the field. The leader may also: (a) add or remove anti-mines from the group for their use (only the anti-mines he or she gives to the group may be used); (b) move the end of the minefield closer to or further from the group; (c) catch people talking or touching the ground and send them back to the start; or (d) forgive those who step off the anti-mine by not sending them back (this option should not be used except when there are extreme time constraints).

The group is usually rewarded (e.g., adding an anti-mine or shortening the field) when it is working together well and progress-

ing toward the goal. It may lose an anti-mine; find the field lengthened; or get blown up when group members talk, laugh, make individual rather than group efforts, or in general are not working together and helping each other. The addition or removal of one anti-mine may seem insignificant, but after the group has been struggling for an hour or more and has become highly involved, one anti-mine can be extremely meaningful.

These are the basic rules of the minefield activity. As facilitators gain experience and confidence, they tend to add personal refinements to optimize the experience for the participants. Some refinements include leaving the anti-mines on the field when the group must start over and imposing additional time constraints.

The Challenge as Applied to Engaged Learning

The Challenge is engaged learning in that it teaches complex principles through the use of structured behavioral activities (Dutton and Stumpf, 1991; Thatcher, 1990). Within the scenario of The Challenge, participants explore new roles and relationships that influence their person-environment relationships with respect to cognitions, conations, and affect. These relationships can be established or changed for a particular individual and a particular set of complex principles. The following section presents the elements that enhance the effectiveness of attitude change and thus are elements of good experiential education as they relate to The Challenge.

A balance of content and process. While the focus is on the processes involved in crossing the minefield, the participants are encouraged to examine the process by looking for the symbols involved in the activity and relating them to the process of opportunity identification, evaluation and execution.

Student rather than teacher based. The leader (teacher) sets the parameters and enforces the rules and allows the students to develop their own strategies and tactics. The teacher may take a more active role through selective reinforcement and punishment to encourage more productive behavior.

Personal not impersonal in nature. The minefield is a stressful situation in which the participants perform unspecified tactics to accomplish specific objectives. It is personal. They must be engaged in the situation or it becomes obvious very quickly that the group will not succeed.

Holistic rather than reductionistic. The minefield is a relatively complex situation in which students can explore ideas and strategies. Participants must think on the go, try multiple tactics and constantly

adjust their actions to accomplish their objective. It not only involves micro adjustments of individual and group behavior but it also requires a macro perspective, thinking in holistic terms to cross the minefield successfully.

Creates an emotional investment. Participants must be engaged or invested physically, emotionally and mentally. Their emotional investment is obvious by their reaction when they fail, and they will fail, as well as when the last person comes off the minefield.

Reflection and self-examination. There must be a debriefing session following the activity in which participants share the experience from their multiple perspectives. In sharing, insights are gained that participants might not have gained otherwise. This discussion can be deep and rich in terms of the content and application of principles that may have been abstract concepts up to that point in time.

Moving outside one's comfort zone: Completing the minefield is not easy. Because it is not easy, it requires focused and concentrated effort to complete the task. It can be very stressful. Often, individual liabilities, particularly physical liabilities become obvious to both the individual and the group and both must learn to compensate for those liabilities. This is particularly true of establishing and maintain balance on rocks or blocks in the minefield.

Conclusion

This paper has provided a conceptual and theoretical framework for understanding experiential education as a rigorous experience-based program of entrepreneurship education. Rigorous means that the educational pedagogy will maintain a high level of excellence in the (a) design of effective experiential activities, (b) effective execution of existing activities, and (c) defense of the use of experiential pedagogy from a theoretical and methodological standpoint against traditional educational approaches.

This paper represents a first step in linking experiential education to the entrepreneurial mindset. Additional empirical research needs to be done to further demonstrate the linkage. The area of experiential entrepreneurship education is still at a relatively early stage of conceptual development. Further research in this area should include categorizing the types of experiential education and the degree to which they engage a student's affective, cognitive, and conative interactions with the environment and enhances their entrepreneurial mindset. This research provides insight into how and why experiential education influences learning in general and the entrepreneurial mindset in particular.

References

Adcroft, A., Wills, A., and Dhaliwal, S. (2004) "Missing the point? Management education and entrepreneurship", *Management Decision*, 42 (3), 512-521

Ajzen, I. (1982) On behaving in accordance with one's attitudes. In M.P. Zanna, E. T. Higgins, and C. P. Herman (Eds.), *Consistency in social behavior: The Ontario symposium* (Vol. 2, pp. 131-146). Hillsdale, NJ: Lawrence Erlbaum.

Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *Handbook of social psychology* (pp. 709-884). Worcester, MA: Clark University.

Bar-Tal, Y. (1992) The effect of the experience with attitude object on the relationships among cognitive and affective components of attitude and behavioral intentions. *Psychological Record*, 42(1), 131-40.

Bennett, R. (2006) "Business lecturers' perceptions of the nature of entrepreneurship", *International Journal of Entrepreneurial Behavior & Research*, 12 (3), 165-188.

Breckler, S. J. (1983) Validation of affect, behavior, and cognition as distinct components of attitude (Vol. 44(11B), 3569): *Dissertation Abstracts International*.

Breckler, S. J. (1984). Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of Personality and Social Psychology*, 47, 1191-1205.

Carlson, S. D. (1985) Consistency of attitude components: A new proposal for an old problem. *Dissertation Abstracts International*, 46(09B), 3261.

Carlson, S. D. (1994) *Entrepreneurship education for the multiple cognitive-developmental levels of third graders: Differential impacts of three experience-based programs*. Los Angeles: University of California.

Chapman, S., McPhee, P., and Proudman, B. (1995) What is Experiential Education, in K. Warren, M. Sakofs, J. S. Hunt, Jr., (eds.), *The Theory of Experiential Education*. Kendall/Hunt Publishing Company, Dubeque Iowa.

Collinson, E., and Quinn, L. (2002) "The impact of collaboration between industry and academia on SME growth", *Journal of Marketing Management*, 18 (3/4), 415-435.

Curran, J., and Stanworth, J. (1989) "Education and training for enterprise: some problems of classification, evaluation, policy and research", *International Small Business Journal*, 7 (2), 11-22.

Crosby, A., (1995) *A Critical Look: The Philosophical Foundations of Experiential Education*. In K. Warren, M. Sakofs, J. S. Hunt, Jr., (eds.), *The Theory of Experiential Education*. Kendall/Hunt Publishing Company, Dubuque Iowa.

Davies, J., Hides, M., and Powell, J., (2002) "Defining the development needs of entrepreneurs of SMEs", *Education + Training*, 44 (8/9), 406-413.

Deamer, I., and Earle, L. (2004) "Searching for entrepreneurship", *Industrial and Commercial Training*, 36 (3), 99-103.

Dutton, J. E., and Stumpf, S. A. (1991) Using behavioral simulations to study strategic processes. *Simulation & Gaming*, 22, 149-173.

Engelen, E. (2002) "How innovative are Dutch immigrant entrepreneurs? Constructing a framework of assessment", *International Journal of Entrepreneurial Behavior & Research*, 8 (1/2), 69-92.

Garavan, T., and O'Cinneide, B. (1994) "Entrepreneurship and education training programmes: a review and evaluation – part 1", *Journal of European Industrial Training*, 18 (8), 3-12.

Gibb, A. (2002). In pursuit of a new 'enterprise' and 'entrepreneurship' paradigm for learning creative destruction, new values, new ways of doing things and new combination of knowledge. *International Journal of Management Reviews*, 4(3), 233-269.

Gorman, G., Hanlon, D. and King, W. (1997) "Some research perspectives on entrepreneurship education, enterprise education and education for small business management: a ten-year literature review", *International Small Business Journal*, 15 (3), 56-77.

Hambrick, D., and D'Aveni, R. (1988) "Large corporate failure as downwards spirals", *Administrative Science Quarterly*, 33 (1), 1-33.

Harris, M. L., Gibson, S. G., and Taylor, S. R. (2007/2008) Examining the impact of small business institute participation on entrepreneurial attitudes *Journal of Small Business Strategy*, 18(2).

Hills, G. (1988) "Variations in university entrepreneurship education: an empirical study of an evolving field", *Journal of Business Venturing*, 3 (2), 109-122

Henderson, R., and Robertson, M. (1999) "Who wants to be an entrepreneur? Young adult attitudes to entrepreneurship as a career", *Education + Training*, 41 (5), 236-245.

Huefner, J. (1991) Affect-cognition-conation consistency and the attitude-behavior relationship. Ph.D. Dissertation, Brigham Young University, Provo.

Hynes, B. (1996) "Entrepreneurship education and training: introducing entrepreneurship into non-business disciplines", *Journal of European Industrial Training*, 20 (8), 10-17.

Ibrahim, A., and Soufani, K. (2002) "Entrepreneurship education and training in Canada: a critical assessment", *Education + training*, 44 (8/9), 421-430.

Jack, S.L., and Anderson, A.R. (1999) "Entrepreneurship education within the enterprise culture: producing reflective practitioners", *International Journal of Entrepreneurial Behavior & Research*, 5 (3), 110-125

Jansen, P., and Van Wees, L. (1994), "Conditions for internal entrepreneurship", *Journal of Management Development*, 13 (9), 34-51.

Joplin, L. (1995). On Defining Experiential Education, in K. Warren, M. Sakofs, J. S. Hunt, Jr. (eds.) *The Theory of Experiential Education*. Kendall/Hunt Publishing Company, Dubuque Iowa.

Katz, D., and Stotland, E. (1959) A preliminary statement to a theory of attitude structure and change. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 3, pp. 423-475). New York: The Free Press.

Keogh, W., and Gallaway, L. (2004) "Teaching enterprise in vocational disciplines: reflecting on positive experience", *Management Decision*, 42 (3/4), 531-542.

Keys, B., and Wolfe, J. (1990) The role of management games and simulations in education research. *Journal of Management*, 16, 307-336.

Klappa, R. (2004) "Government goals and entrepreneurship education", *Education + Training*, 46 (3), 127-138

Kothandapani, V. (1971) Validation of feeling, belief, and intention to act as three components of attitude and their contribution to prediction of contraceptive behavior. *Journal of Personality and Social Psychology*, 19, 321 - 333

Kristensen, K. B., Pedersen, D. M., and Williams, R. N. (2001) Profiling religious maturity: The relationship of religious attitude components to religious orientations. *Journal for the Scientific Study of Religion*, 40(1), 75-87.

Ladzani, W., and Van Vuuren, J. (2002) "Entrepreneurship training for emerging SMEs in South Africa", *Journal of Small Business Management*, 40 (2), 154-162.

McGuire, W. J. (1969) The nature of attitudes and attitude change. In G. Lindzey and E. Aronson (Eds.), *The handbook of social psychology* (2nd ed., pp. 136-314). Reading, MA: Addison-Wesley

McMullan, C. A., and Boberg, A. L. (1991) The relative effectiveness of projects in teaching entrepreneurship. *Journal of Small Business and Entrepreneurship*, 9, 14-24.

Mitchell, T.R., and James, L.R. (1989) Conclusions and future directions. *Academy of Management Review* [Spec. iss.], 14, 401-408

Morris, M., Kuratko, D. and Schindehutte, M. (2001) "Towards integration: understanding integration through frameworks", *International Journal of Entrepreneurship and Innovation*, 2 (1), 35-49

Ostrom, T. M. (1969) The relationship between the affective, behavioral, and cognitive components of attitude. *Journal of Experimental Social Psychology*, 5, 12 - 30.

Peterman, N., and Kennedy, J. (2003) "Enterprise education: influencing students' perceptions of entrepreneurship", *Entrepreneurship Theory and Practice*, 28 (2), 129-145.

Rasheed, H. S., and Rasheed, B. Y. (2003) Developing entrepreneurial characteristics in minority Youth: The effects of education and enterprise experience. *International Research in the Business Disciplines*, 4, 261-277.

Robinson, P. B. (1987). Prediction of Entrepreneurship Based on an Attitude Consistency Model. Unpublished Doctoral Dissertation, Brigham Young University, Provo.

Robinson, P. B. (1996) The MINEFIELD exercise: 'The challenge' in entrepreneurship education. *Simulation & Gaming*, Vol. 27 (3), p. 350-365

Robinson, P. B. (2010) Engaged Learning and the Entrepreneurial Mindset, Presented at the Utah Academy of Science, Arts, and Letters annual meeting, St. George, UT.

Robinson, P. B. and Malach, S. (2004) Multi-Disciplinary Entrepreneurship Clinic: Experiential Education In Theory and Practice. *Journal of Small Business and Entrepreneurship*, Vol. 17(4), pp 317-332

Robinson, P. B., Stimpson, D. V., Huffner, J. C., and Hunt, H. K. (1991) An attitude approach to the prediction of entrepreneurship. *Entrepreneurship: Theory and Practice*, 15(4), 13-31.

Rosenberg, M. J, and Hovland, C. I. (1960) Cognitive, affective and behavioral components of attitudes. In M. J. Rosenberg, C. I. Hovland, W. J. McGuire, R. P. Abelson, and J. W. Brehm (Eds.), *Attitude organization and change: An analysis of consistency among attitude components*. New Haven, CT: Yale University Press.

Sexton, D., and Bowman, N. (1984) "Entrepreneurship education: Suggestions for increasing effectiveness", *Journal of Small Business Management*, 22 (2), 18-25.

Shane, S., and Venkataraman, S. (2000) The promise of entrepreneurship as a field of research. *Academy of Management Review*, 25, 217 - 226.

Shaver, K. G. (1987) *Principles of social psychology*. Cambridge, MA: Winthrop.

Souitaris, v., Zerbinati, S., and Al-Laham, A. (2007) Do entrepreneurship programmes raise entrepreneurial intention of science and engineering students? The effect of learning, inspiration and resources. *Journal of Business Venturing* 22 (4), 566-591

Smith, M. B. (1947) The personal setting of public opinion: A study of attitudes toward Russia. *Public Opinion Quarterly*, 11, 507-523.

Specht, L. B., and Sandlin, P. K. (1991) The differential effects of experiential learning activities and traditional lecture classes in accounting. *Simulation & Gaming*, 22, 196-210.

Tam, H. (2009) How and to what extent does entrepreneurship education make students more entrepreneurial? A California case of the Technology Management Program. Santa Barbara: University of California, Santa Barbara.

Tam, H., Hansen, G., Blomstrom, S., and Robinson, P. B. (2009, June 14 - 17) Entrepreneurship Program Assessment by Student Outcome. Paper presented at the Proceedings of the American Society for Engineering Education 116th Annual Conference & Exposition, Austin.

Thatcher, D.C. (1990) Promoting learning through games and simulations. *Simulation & Gaming*, 21, 262-273.

Verplanken, B., Hofstee, G., and Janssen, H, J. (1998) Accessibility of affective versus cognitive components of attitude. *European Journal of Social Psychology*, 28, 23-35.

Practical Theory: A Teacher's Guide to Making Research Writing Theory Useful for Students

Angie McKinnon Carter

Utah Valley University

Abstract

Based on Doug Brent's and Margaret Kantz's claims that rhetorical reading forms the basis for argumentative writing, this paper argues that for students to most effectively compose their research papers the practical aspects of doing research must be connected with the theoretical underpinnings of why we research. I offer specific activities teachers can use to connect theory with practice. For example, although I discuss the importance of knowledge, social interaction, and timing, I stress the importance of having a theoretical framework for classroom activities to make the "nuts and bolts" of research writing actually useful for students. Along the way, I offer specific suggestions on how the theoretical can be made more visible to students through the practice their teachers provide. In my experience, both as a researcher and a teacher, I find that the more practical the activities I provide in class, the more successful my students are with writing their research papers on their own.

Requiring research-based writing is standard in many college and university classes, and many articles have been written discussing how teachers can present and teach research papers. In *Reading as Rhetorical Invention: Knowledge, Persuasion, and the Teaching of Research-Based Writing*, Doug Brent asserts that much of the problem with articles about teaching research writing is that they are lists of suggestions dealing with the “superficial aspects of the research paper” such as using note cards correctly or avoiding plagiarism (103). He argues that what we need more than a list of instructions on how to teach the research paper is “an encompassing definition of what it really means to compose discourse based on other people’s texts” (103). Indeed, much of what he advocates telling students about research papers is what composition teachers, as well as teachers from other disciplines, already know about research. Briefly, research is not data retrieval but a conversation about the data (109); all sources are biased (109); instincts readers have about a text can be interpreted as ethos or pathos and can be used as convincing evidence (109); and research is recursive by nature (109). However, although we (teachers across the university and composition instructors in particular) attempt to teach these ideas to our students, as Margaret Kantz explains in “Helping Students Use Textual Sources Persuasively,” our students tend to get bogged down in the practicalities of finding sources, summarizing them, and informing us (their teachers) of what they have learned.

In our minds, we have a rich, rhetorical view of reading and writing that we teach to our students. Yet somehow the theory gets lost when teaching the practical aspects of doing a research paper—finding a topic, taking notes, writing the paper, etc. Therefore, I argue that for students to most effectively compose their research papers the practical aspects of doing research must be connected with the theoretical underpinnings of why people research. In other words, theory (specifically, rhetorical theory) needs to be an explicit rather than implicit part of our teaching. Essentially, we should not only tell students what they should do but explain why they should do it and how their rhetorical goals can contribute to what they do. Thus, in this article, I offer hands-on activities and examples that connect theory with practice. By presenting some ways Brent’s model of rhetorical reading and Kantz’s heuristics using rhetorical gaps can be used, I illustrate how practical theory (practical applications of rhetorical theory) combined with meeting students’ interrelated needs of knowledge, practice, social interaction, and time enables students to produce more engaging and well-researched papers.

Students Need Knowledge

Students need knowledge to be able to write good research papers: topical knowledge and procedural knowledge. That is, they need background information about the subjects and issues that they choose to write about, and they need instruction on how to research and write the paper and practice using those instructions.

Topical Knowledge

Topical knowledge is usually regarded as disciplinary knowledge. As the case studies of my thesis (Carter) illustrate, advanced writing students bring considerable disciplinary knowledge to their topic choices. However, although less-experienced writing students lack disciplinary knowledge, I have noticed that they often have experiential and intuitive knowledge (or personal knowledge) that they can apply to disciplinary issues. Furthermore, as Brent argues, disciplinary knowledge is actually formed through active inquiry (107) and by interacting with other worldviews (106). Thus, when students can connect disciplinary knowledge with personal knowledge, they actually do better research. In fact, I have noticed, both when conducting my thesis study and while teaching, that students tend to write better papers when they offer an original argument to address a problem or question that they have identified. Essentially, they view their research papers, as Kantz says, “as an opportunity to teach someone, to solve someone’s problem, or to answer someone’s question” (83). This is an audience-focused way (and, indeed, a more rhetorical way) of finding a research topic.

Finding a topic of personal interest worked very well for Jeremy,¹ a student in my English 2020 class (Intermediate Writing for Sciences and Technology). For his final researched argument paper, his initial topic concerned the implications of the Navajo Nation’s recent vote to allow casino gaming on the Navajo reservation—his home. Before he had started doing any library research, he knew about the vote, and he had heard that casino gaming could be a problem. He knew about his own tribe’s culture, living conditions, and traditions. He also was aware of many common stereotypes about Native Americans. This personal knowledge provided some background knowledge that allowed him to connect with sources he found.

¹ Jeremy is a pseudonym as this student prefers that his real name not be used.

Procedural Knowledge

Jeremy's personal knowledge was not yet useful, however, because he did not know how to leverage it to construct a research paper. Jeremy's experience echoes Brent's contention that students already have complex knowledge structures based on their life experiences, but they frequently do not know how to tap into them and use them (104, 106, 111-12). We can help students critically examine their experiences along with their texts by helping them understand how their perceptions of the world influence how they make sense of other people's ideas. Essentially, we need to teach them how their worldviews interact with other people's ideas. Thus, my second point regarding students needing knowledge is that, in addition to topical knowledge, students need procedural knowledge.² For me this involves spending some time early each semester presenting Kinneavy's rhetorical triangle and using it to help students find gaps in texts (Kantz 83).

Rhetorical Theory and Rhetorical Gaps

When I teach Kinneavy's rhetorical triangle (Figure 1), I also teach Kantz's heuristic for using rhetorical gaps (Figure 2), and a basic understanding of both is needed to understand Jeremy's experience of pulling his paper together. As Kantz explains, the three points of the rhetorical triangle are always present in a piece of writing no matter how, or even whether, the writer considers them (80). An author always constructs a text to address a particular situation and/or topic for a particular audience. However, Kantz goes beyond looking at each point individually and asks students to consider how the points interact. For instance, the ways in which writers consider their audience will affect how they present the subject to that audience (81). Examining the interaction between two of the rhetorical triangle's points provides a starting point to consider how texts are constructed. As students learn what should happen in a text, we can then teach them to identify places where what is supposed to happen does not, what Kantz refers to as a gap (81). For instance, although an author can write to one audience (e.g., psychologists), someone from a different group (e.g., students learning about eating disorders) can also read that text (81). If the intended audience and the actual audience are different enough, a gap may form that can create misunderstandings or alternate views.

² By procedural knowledge, I mean both what students know about how to research and write a paper and what they actually do when they research and write.

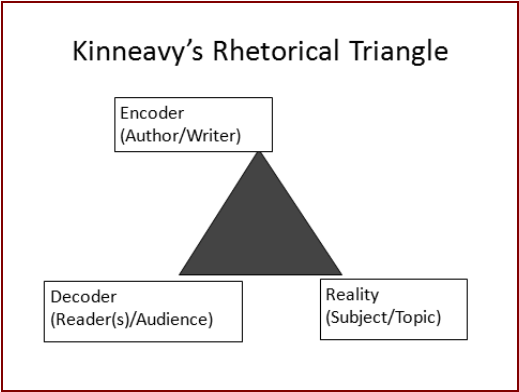


Figure 1: A graphic representation of Kinneavy's rhetorical triangle created from Kantz's verbal description of the same (79-80).

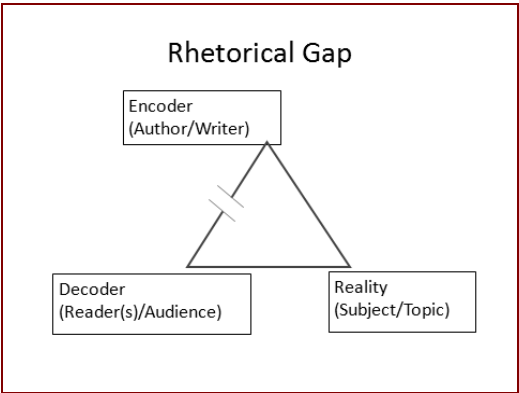


Figure 2: Rhetorical gap illustration created from Kantz's verbal description (83).

Jeremy's research-writing experiences illustrate this concept. Even though Jeremy had rich personal knowledge and a strong connection to his sources, he struggled with how to present the information he was finding. In his research, he uncovered some startling statistics indicating that tribes on the East Coast that had embraced casino gaming many years ago were now experiencing serious social ills and had start-

ed to lose aspects of their traditional culture, even as they reaped the promised economic prosperity. He was concerned that if the Navajo Nation adopted casino gaming the Navajo people would lose some key attributes of their culture, and he wondered whether the money was really worth the loss of his heritage. While he had to do research to discover the effects casino gaming had had on other tribes, as previously mentioned, he already had considerable personal knowledge about his tribe. Although he did not realize it at first, he was identifying a subject–audience gap, a gap between what his audience thought the results of gaming were (economic prosperity) and some unintended results that had not been foreseen (the loss of Native American cultures).

During a conference we had on the issue, I suggested that he consider his tribal elders, the people who would ultimately make the decisions about casino gaming for the tribe, as his audience. Where in a previous draft he had struggled to present his ideas, his final draft was much improved. Because he had a real audience to accompany the real problem that he had already identified (essentially connecting all the points of the triangle instead of leaving one edge open), his writing became fresh and insightful. Also, because he had identified a gap in how to perceive the economic benefits of gaming, he had an issue to discuss. He ended up arguing that the tribe should proceed with casino gaming with caution to avoid creating the societal problems that had plagued other tribes and to ensure that their unique tribal heritage would not be lost. By focusing on a specific audience (the tribal elders) and addressing his concern that they did not have some pertinent information, Jeremy was not just “trying to solve [the tribal elders’] problem” (Kantz 83), he was trying to prevent them from creating one.

Although we had discussed the rhetorical triangle and the concept of gaps in class, it was not until Jeremy and I made the theory applicable to his particular situation that he was able to actually use the theoretical principles to improve his writing. By helping him picture a real audience and formulate a topic and an approach to it that would interest them, he created a document that had real value to him and to his community. What I believe happened in Jeremy’s case was his topical knowledge provided an entrance to the sources. Then, realizing the conversational nature of reading and writing and identifying a gap (a la Kantz’s heuristic) between the promise of casino gaming on Native American reservations (his audience’s view of the subject) compared with the potential for losing his culture (the topic he had discovered in his research) gave him an avenue to present an original idea. In sum, he needed some procedural knowledge for interacting with his sources,

which led him to realize that he had an original argument to contribute to the conversation.

Students Need Practice

Jeremy's experience also suggests students' need for practice. He had to work with his sources and draft several times to develop his ideas. In addition to having students work on their own, providing class activities that allow students to practice what they have been taught, what Brent calls "constructing activities" (108), is also valuable.

Pre-Draft Assignments

Once students have a research topic and question, I find that students need to practice applying the concepts we are discussing in class before writing a full rough draft. As Brent notes, "[I]nstruction on the research process . . . typically . . . deals with the beginning and the end of the process (using the library and writing the drafts), but it has a gaping hole in the middle where much of the real work of knowledge construction is performed" (105). I use several intermediate assignments such as checking students' notes, having them bring summaries of their sources to class, bringing a single paragraph of analysis, and so on to help students bridge the gap between finding a research topic and writing the first draft. Step by step, as students work with a single source or small portions of their paper at a time, they generally find that by the time it is time to construct a full rough draft much of the work has already been done. Moreover, by working on the draft incrementally, students learn to respond to their sources rather than just collate them. By focusing on this "writing in the middle,"³ we can focus our time and energy on developing the thinking processes that we want to see illustrated in their drafts.

Unfortunately, students often do not recognize our, or their, pedagogical reasons for these assignments. As my students work on these pre-draft assignments, I have learned that it is extremely important to remind students about these assignments' value and to connect them with the rhetorical principles we have learned previously in the semester. For instance, I require my students to take notes on their sources and submit these for review. I am looking for ways that they are interacting with their texts—making comments in the margins, asking questions, and so on. If their note-taking consists mainly of highlighting,

³ This is an expression I picked up from an assignment sheet Doug Downs created. Downs is currently an assistant professor at the University of Montana.

then chances are they are viewing their sources as information sources rather than as arguments (Kantz 79). This lets me know that I need to do some re-teaching to remind students that the rhetorical analysis we did earlier in the semester still applies as they do research. In other words, they need to recognize what worldviews are being presented in their texts and how they are being persuaded by them (Brent 105). Once they have done this work, they can formulate how their worldviews interact with their texts' worldviews and create an original working thesis. I also find that asking students to submit a preliminary working thesis before they write a draft saves them time in the long run. If they are not connecting their claim at least implicitly with prior research, I know that they will have a hard time meeting a key requirement for the researched argument paper as I assign it. Essentially, in these pre-draft assignments, I help students discover the conversational nature of writing: that is, using sources to have a conversation about their research question.

Locating Credible Sources

One area where students need a lot of practice is locating and then knowing how to read credible sources. I have noticed that many of my students have trouble distinguishing a scholarly source from a more popular source. To help them recognize the differences, I conduct what I call an Audience, Purpose, Genre, Structure, and Style (APGSS) Workshop.⁴ Its purpose is for students to deductively identify different patterns evidenced in scholarly and non-scholarly articles and to realize how audience considerations shape genre. The entire exercise is rooted in using the rhetorical triangle (author/reader/topic) to determine how the ways in which these elements intersect influence the type of writing that is produced. It is important for students to do the workshop and connect it to the theory that I have previously presented in class. Without this connection, many students walk away from the activity considering it little more than busywork.

The activity follows this general pattern. When possible, I take my students to the Current Periodicals section of the library.⁵ I prefer

⁴ I adapted this exercise from one described in a "For Writing and Discussion" exercise (Ramage et al. 55).

⁵ When it's not possible, I have been able to check out four or five sets (eight to ten periodicals total) to take to class. While this works for the activity, I prefer taking my students to the library, so they get experience being in the library, finding call numbers, and just being in that atmosphere of quiet, purposeful study.

using the print versions of journals and magazines for this activity because seeing the periodical in its entirety provides more context for how the entire source is geared toward a specific audience's needs. Later, when students access electronic sources in the databases, remembering the contextual clues from the print sources can help them better determine whether an electronic source is scholarly or not. Prior to class, I print a worksheet listing areas students should evaluate in the two periodicals (one scholarly source and one popular source) including the following: the number and types of ads, the length of the articles, the length and style of the sentences, the type of vocabulary used, the authors and their backgrounds, the publisher of the source as well as other criteria. Students select two periodicals, a scholarly source and a popular source, from paired lists I have included on the APGSS worksheet (e.g., I might pair *American Political Science Review* with *Time*). After students have compiled the information above about the two sources, I ask them to record responses to the following questions, which we later discuss as a class:

- What features does a scholarly text have? How can these features help you distinguish a scholarly source from a popular one?
- How does the audience an author believes s/he is writing to affect what and how s/he writes? How does this idea manifest itself in the differences between scholarly journals and popular magazines?

Notice that these questions are rooted in understanding how the connections among the points of the rhetorical triangle create different types of writing. Discussing these questions with students after they have completed the workshop helps them make the connection that the way a text is created is influenced by the intersection of how the writer perceives the audience's views of the subject matter.

The benefits of this exercise are numerous. First, because students have spent time reviewing two sources that present similar subject matter side by side, they start to realize that who authors consider their readers to be affects not only what is written but how it is written. Second, they realize that they often are not the intended audience for much of the scholarly articles they will be asked to read. Because we have discussed Kantz's concept of rhetorical gaps previously in the semester, it is easier to help students realize that they are not the intended audience. Without this audience-focused look at scholarly texts, students often see scholarly texts as being unnecessarily dense. To focus their attention while doing research, I assign them to keep research logs, which encourage them to read their sources more slowly, reflect on

what they read, and be more willing to reread their sources. Because students are encouraged to read as a participant in an academic conversation, they tend to be less critical of the authors who wrote the texts that they are reading (including faulting the authors' writing styles as boring). Once they realize that they are infiltrating new territory, many students discover that scholarly writers offer richer arguments and better data than easier-to-access sources.

Highlighting Activity (Conversing with Sources)

Another area in which students need a lot of practice in seeing how their choices affect their readers is in how to have a conversation with their sources. In other words, they need some practice in seeing how well they are balancing their own ideas with those of their sources. The following activity provides a visual way for students to see how much interaction they are having with their sources. For this activity, students need at least a partial rough draft of a paper in which they use sources. If I do this activity in a computer lab, students need electronic access to their draft (on a flash drive or via email), so they can access it in the lab. If we do this activity in a classroom, they need at least four different colors of colored pencils or highlighters. Before having students work on their own papers, I show a former student's paper in the same genre we are working on.⁶ I write the following four categories on the board and assign each a color: topic sentence (blue), evidence from a source (green), commentary (yellow), and transitions (red). I post the model on the screen, and as a class we identify the topic sentence, evidence from a source, commentary, and transitions. As students identify each portion of the paragraph, I highlight the sentence (or portion of each sentence) in the color that corresponds to each category.

After color-coding the paragraph, we discuss the patterns that emerge. For instance, if a paragraph is mostly green with only one sentence of yellow, then we can determine that the student did not have a topic sentence and the evidence from the text is possibly taking up too much of this part of the paper. The lack of red also indicates that this paragraph is not strongly connected to the rest of the paper. If the student has all four colors, then we can determine if the order of the sentences is appropriate for what the student is trying to accomplish in this paragraph. For instance, the location of the topic sentence (blue) may vary depending on the author's needs and the genre of the paper. Stu-

⁶ For instance, if we are working on the researched argument paper, I would show a former student's researched argument paper as a model.

dents are generally very good at pointing out whether a certain organizational structure works or not, and seeing the color-coding enables them to better discuss why they are having a certain reaction to a paper.

After we finish working with the model, students do this exercise with the paper they have brought. Once they have identified and coded their sentences, they can discuss with a partner what they have learned. For homework (or if they have time in class), I ask them to revise their paragraphs based on what they have discovered. For instance, if they have too much evidence from their sources, then they might insert commentary strategically throughout the paragraph to explain the value of that evidence. If they have too much commentary (another common problem), I ask them to find specific passages in their sources that relate to their discussion that they can add to their paper. Often students are implicitly responding to sources, but they just have not included specifics from a source so the reader does not see where the student writer is coming from. In essence, the first part of the conversation is missing. (It is like listening to one half of a telephone conversation.) In sum, reminding students that they are conversing with their sources—that the authors of their sources could be considered part of their audience (Greene and Lidinsky 167)—could help them approach the task of using their sources in a more rhetorical manner.

Students Need Social Interaction

A key aspect of Brent's model of rhetorical reading, to which many composition instructors already seem to subscribe, is that knowledge is socially constructed.⁷ That is, people make knowledge by interacting with others whether this happens through speech or writing. The next step, then, requires students to see us and their peers as parts of their audience—a ready-made venue for conversations about their texts to occur. Brent advocates providing opportunities for students to do more interacting to improve their research papers.

Brent strongly recommends that students collaborate in one form or another. Although he advocates using the collaborative research paper, I generally find it impractical in the introductory-level writing courses I teach. However, this does not mean that collaboration on the

⁷ The history of how this approach became widely accepted in composition theory is extensive and beyond the scope of this paper to review. For a brief review of how the field of composition has adopted elements of a social-constructionist approach that itself emerged from epistemic rhetoric, see Richard McNabb's article in *Theorizing Composition: A Critical Sourcebook of Theory and Scholarship in Contemporary Composition Studies* (105–106).

research paper is impossible. On the contrary, teachers can be doing more in this area. Currently, most classroom collaboration for research writing consists of having students meet for peer response groups and critique each other's rough drafts. Unfortunately, this comes at the end of the process. Students need feedback throughout the process, especially when reading and responding to their sources. Having more interaction at the beginning and middle of the process may improve the quality of students' final papers and the quality of the critiques they give each other's drafts at the end of the process. Moreover, simply sharing their research, or even quizzing each other about it, can force them to articulate what they are learning as well as make insights and connections with their research they might not have made while reading on their own.

Modeling with Student Papers

Earlier, I showed that students frequently talk about writing in my classes. As I demonstrated in the previous section, I use student papers to illustrate how revision (substantive change) can improve papers. As students analyze previous students' papers for various features, their intuitive knowledge about how to present research becomes explicit. For example, they can identify features that make an analysis different from a summary, or they can explore the attitude authors convey with their tone and the effects this has on their writing. Furthermore, as students work together, they learn ideas from each other that they would not have had otherwise. By having these class critiquing sessions, other teachers' and my experience has confirmed that students are able to make more thoughtful and helpful comments about their peers' papers during peer work-shopping sessions (Oates).⁸

Over the years as I have used models in my classes, I have discovered, surprisingly, that my students like looking at models that do not completely meet the requirements. By having a reader's experience in reading a draft that misses the mark, they learn how not meeting the standard for that paper (generally, as presented in a rubric) can negatively impact their readers' ability to read their work as intended, making them more motivated to meet the standard. Even if students still struggle to meet it, however, they come away from these sessions un-

⁸ I was originally introduced to the idea of using student papers for modeling by Jared Oates, at the time a second-year graduate student and graduate instructor for both freshman composition and advanced writing. He originally shared this insight with me when we were both graduate instructors at Brigham Young University in the late 1990s.

derstanding the concept of audience better as well as seeing their role as a writer in making the reader's experience more enjoyable. Moreover, by seeing the ways in which a fellow student met the requirements in some areas but fell short in others, they gain confidence in trying to meet the requirements in their own papers.

Group Conferencing

Another method I use to encourage more feedback during the drafting process is group conferencing. As William Perry indicates, the impetus to change is frequently influenced by interactions with peers first (69), so I try to provide many opportunities for students to consult with each other. Toward the end of the research project, group conferencing functions like a regular peer-review session: Students read each other's papers and comment on them. However, the group conference has some important differences. For instance, before drafting the papers, these conference groups can be used as research groups. Essentially, students can consult with members of the group about their papers throughout the process. In fact, each class period includes some time for students to meet in their groups and talk about their papers (a need I discuss in the next section).

Once students have produced an initial draft, another key difference between the group conference and a regular peer-review session emerges. I ask students to give each other their papers between five days to a week before we actually have the conferences. This time frame allows students to carefully read each group member's paper and respond to it.⁹ I either give students a detailed list of questions to address or ask them to use the previously provided rubric to respond to their peers' drafts. Students are required to post their responses to each group members' paper two days prior to the first conference session. Before coming to the conference, they are then required to have read their group members' comments about their paper. The conference session then becomes a place to have a conversation about each person's draft. Since everyone has responded to the initial draft and read the initial responses about it, the conference becomes a place to receive clarification on ideas or, perhaps, bounce ideas for revision off the peer-review panel.

⁹ I use a Learning Management System such as Blackboard Vista to set up a discussion group to facilitate the distribution of papers and the initial response to those papers.

As a teacher, I facilitate these conferences in a couple of ways. First, I generally assign the groups. If student schedules permit, I try to group students who are writing on a similar issue. This allows a sense of community to develop, which provides a more authentic audience since people tend to read material in which they have some interest. Second, we meet in my office to discuss each group's papers. Because I have more drafts than I am able to read in a three-day period, I often do not have my first reading of their papers finished until the students arrive for the conference. This means they have read each others' initial response to their papers but not mine before they come to the conference. This has rarely been a problem, though. In fact, it has provided an unexpected benefit of allowing the students to share ideas with each other more freely. That is, they do not worry if they have read their peers' papers "wrong" if I happen to disagree. I have been pleased with the discussions that develop online, sometimes spontaneously, as students converse back and forth with each other about their drafts. When we do meet for the face-to-face conference, more often than not, I find that I often concur with students' evaluations, allowing them to feel more confident in their ability to notice both problems and excellence. As a result, I believe that real conversation develops—students see each other as valuable voices in offering feedback. Since students now have a real audience, the perennial concern of students only writing to please the teacher diminishes. Overall, the group conference seems to create a real learning community.

I recently had a student, Alyssa,¹⁰ whose experience provides a powerful illustration of the benefits of the group conference. About three weeks before the end of the semester, Alyssa submitted a partial rough draft of the researched argument paper (a draft consisting of about five pages of what would eventually become a 10- to 15-page paper) for a group conference on teenage runaways. Her partner in the group conference, Rachelle,¹¹ was an especially conscientious peer reviewer. Because of their dedication to reviewing each other's papers, we had a lively discussion about their papers. As we reviewed Alyssa's, we noted that she wanted the federal government to create a national system of runaway shelters; however, she had not addressed how the government would pay for these structures or even whether they would have the political will to do so in this age of economic austerity and budget cutting. Rachelle and I suggested that she address this counter-argument in her paper.

¹⁰ Alyssa has asked that she be identified by only her first name.

¹¹ Rachelle, Alyssa's partner, also wishes to be mentioned by her real name.

Alyssa did so in subsequent drafts. Where her first draft only addressed the need for a nationwide system of runaway shelters, in subsequent drafts she addressed how this need could be addressed with no additional cost to the federal government. She only needed about two additional paragraphs to address a significant counterargument. The fact that she was willing to take the suggestion offered in her group conference and incorporate it into her paper illustrates the potential of the group conference approach. Including these additional paragraphs also demonstrated an understanding of her audience. She realized that if she really wanted a nationwide system of runaway shelters, she needed to include in her audience the people who hold the purse-strings for making that goal a reality.

Students Need Time

As Jeremy's and Alyssa's experiences and the writing activities I have described demonstrate, writing a research paper is a time-intensive process. What is often overlooked is that research requires time off-task as well as time on-task. Time off-task is particularly critical because students need to disengage from the process to allow their brains the time to consider what they have already produced. Aaron, one of the participants in my thesis study, noticed that in these "idle" moments or times when the brain works on other tasks, it devises new approaches to the material: "I find that . . . it's usually when I least expect it I find the insight that is important. It's usually when I'm concentrating on something else" (Carter 247). These moments of inspiration create the impetus that guides the next wave of on-task activity, whether that burst of creativity takes the form of a new question to ask the sources or provides the wording for a section of the draft. Therefore, the timing I describe below should give students enough time both on-task and off-task for them to figure out what they want to say and then have time to write it.

Drafting and Revising

As we all know, drafting takes time, but students often have trouble using their drafting time wisely because they do not know what to say about their sources. The pre-drafting assignments that I mentioned previously are crucial in helping students generate ideas that they can use when drafting. If students have done the pre-drafting assignments conscientiously, they should have plenty to write.

Students also need time to produce multiple drafts (I ask for at least two drafts of each paper), and Kantz offers an implicit rationale for why students need multiple drafts and more time between them. She

suggests that students can only handle a few of the rhetorical tasks that we are asking them to juggle at a time (85). Moreover, she continues, most students just need to write down what they know before they create an argument out of it (86). Added to these practical matters, many students have an attitudinal problem in that they expect to produce a great paper in a single draft and believe they are somehow inept if they cannot do it (86). Kantz's perceptions of students' writing-related thought patterns mirror the interactions that I have had with students. Therefore, giving students more time between drafts can allow them time to determine what they want to say about the information that they have discovered. Inevitably, our semesters are crammed with information to present and papers for our students to write, but we still need to create a time frame in which students have some downtime between papers to think about how to proceed next in order to craft papers that meet their audiences' needs. I have found that giving students at least four days between drafts produces better results than asking for a new draft by the next class period (which is often only two days away). In addition, I try to read one draft (or a partial draft) of my students' papers before they turn that paper in for a grade.¹² Students need to receive some audience feedback about their papers. Between the peer reviews and my reviews, they get many reactions from people who are reading their essays.

Also, the best revision comes when enough time has elapsed for the reader to see the draft with fresh eyes. Since it generally takes me at least a couple of weeks to read and respond to two classes' worth of papers, by the time my students receive their papers back, enough time has passed that students are, first, willing to look at the paper again, and second, able to actually see the truth in both the criticism and the compliments. With another five to seven days to consider my comments and re-craft the paper, they have time to rework passages to better meet their purpose for their intended audience. Essentially, with this timing, they get some time off-task between drafts where they are not consciously thinking about the paper, which allows the brain to work on it subconsciously. After the paper is returned, students then have enough time to produce the necessary revisions. In sum, having some procedural knowledge about how to revise (as I discuss in the procedural knowledge section earlier in this paper) combined with the timing I describe in this section should help students produce substantially im-

¹² Doug Downs has used this approach in his classes, and I learned about the idea from him.

proved drafts which should ultimately lead to better quality writing overall.

Conclusion

While I have written this article from the perspective and experience of a composition instructor, one who only teaches writing, I have written not only to my fellow composition teachers but also to teachers of many other disciplines. While the composition classroom may seem the most conducive venue to teach the practical theory that I have presented, students in classrooms across the range of disciplines could benefit from having their teachers apply these strategies to their classes. The ways in which the points of the rhetorical triangle are applied and intersect to create texts vary from discipline to discipline. Across the curriculum, teachers clearly know the disciplinary knowledge of their fields, but they also know, intuitively and often explicitly, the rhetorical knowledge needed to make that disciplinary knowledge accessible to others in their fields. Thus, a business professor probably understands the intricacies of reaching business audiences better than a composition professor. Likewise, physicists understand how to craft documents for other physicists, so they are best qualified to teach their students how to match physics topics to interested audiences. While not every activity I have described would work in every class, several could be adapted to meet various instructors' needs. Students need their instructors to share with them how knowledge is created and presented in the entire range of disciplines. The more that practical theory can be used by instructors from a variety of disciplines, the more likely it is that students can maintain the gains that they make in the writing classroom, enabling them to be effective researchers and writers long after they have taken their last writing class.

Works Cited

Brent, Doug. *Reading as Rhetorical Invention: Knowledge, Persuasion, and the Teaching of Research-Based Writing*. Urbana: NCTE, 1992. Print.

Carter, Angie McKinnon. "Improving Students' Research and Report Writing Skills: A Comparative Analysis of Freshmen and Upperclassmen's Composing Processes for Research Reports." MA thesis. Brigham Young U, 1999. Print.

Greene, Stuart, and April Lidinsky. *From Inquiry to Academic Writing: A Text and Reader*. Boston: Bedford/St. Martin's, 2008. Print.

Kantz, Margaret. "Helping Students Use Textual Sources Persuasively." *College English* 52.1 (1990): 74-91. JSTOR. Web. 9 Aug. 2007.

McNabb, Richard. "Epistemic Rhetoric and Theories." *Theorizing Composition: A Critical Sourcebook of Theory and Scholarship in Contemporary Composition Studies*. Ed. Mary Lynch Kennedy. United States: Mary Lynch Kennedy, 1998.

Oates, Jared. Personal Interview. 26 Jan. 1999.

Perry, William. *Forms of Intellectual and Ethical Development in the College Years: A Scheme*. New York: Holt, Rinehart & Winston, 1970. Print.

Ramage, John D., John C. Bean, and June Johnson. *The Allyn & Bacon Guide to Writing*. 4th ed. New York: Pearson/Longman, 2006. Print.

Developing Instruction for the Diverse Classroom: Collegial Collaboration in Action

Vessela Ilieva

Utah Valley University

Abstract

The number of ethno-culturally and linguistically diverse students in Utah and across the United States is increasing faster than any predictions. Our schools struggle to respond adequately to this demographic change. This research reflects on one approach to presenting mathematics teachers with opportunities to acquire a variety of knowledge and skills needed to teach effectively all students. It explores the involvement of an English as a Second Language (ESL) professional with a team of mathematics teachers to collaboratively plan and review lessons to successfully support the academic learning of ethnically and linguistically diverse students in the mainstream content classroom. The ESL teacher helped identify needed adjustment in instruction for students who have not yet developed sufficient linguistic or cultural knowledge needed to make sense of academic tasks and mathematical problems as originally envisioned by the mathematics teachers. The ESL educator further clarified how these missing connections might prevent students from grasping important mathematical knowledge and

contributed to engaging all students in meaningful learning that considered their different backgrounds as part of the teaching process.

Recently published census data confirmed that the United States continues to grow as an increasingly diverse nation (U. S. Census Bureau, 2011). Today in Utah, one in every four school-aged children is a member of a group considered an ethnic or racial minority. A substantial part of the current educational discourse is devoted to identifying ways in which schools can best provide these diverse students with quality educational experiences; however, the resulting educational change, if achieved, is occurring significantly more slowly than the changes in student body makeup. One consequence of this discrepancy is the documented lower academic achievement of diverse students compared with their mainstream peers in core content areas. In mathematics, the National Council of Teachers of Mathematics (NCTM) defined this achievement gap as an indicator of significant differences in achievement between groups of students based on race, ethnicity, or language, among other factors (NCTM, 2004). Despite an overall increase in achievement of all students in the last decades, the gap has remained, and minority group membership is largely associated with notably lower mathematics achievement (Nasir and Cobb, 2002).

Purpose of the Study

This study explored one model of engaging teachers in professional conversations geared toward improving the opportunities for learning mathematics for diverse students. The research sought to find ways in which English as a Second Language (ESL) teachers could support mathematics teachers in their efforts to change and improve instructional practices and provide quality learning experiences to ethno-culturally and linguistically diverse students. Mathematics is generally considered a universal, culture-free subject that uses an abstract and common language, largely in the form of representations and conventions (Nasir et al., 2008). As a result, there exists an erroneous expectation that learners new to a school system should demonstrate knowledge of and learn mathematics regardless of their current level of proficiency in academic English or if they have previously learned mathematics in another country. In fact, the learning of valuable mathematics is culture dependent, and thus gaining proficiency in mathematics presents challenges not always anticipated or acknowledged (Nasir et al., 2008). In the focal point of any effort to provide appropriate opportunities for all mathematics students to learn and achieve are

their teachers (Nasir et al., 2011). These professionals often lack sufficient training and skills in teaching culturally and linguistically diverse students, and collaboration with an experienced ESL teacher could provide them the support they need to successfully teach mathematics to all students.

Review of the Literature

This study was designed on the intersection of two frameworks: culturally responsive instruction and theories of teaching and learning non-native languages. The exploration further built on available scholarly work focused on the need for and potential of collaborations between ESL specialists and content-area teachers. The specific form of teacher collaboration utilized in this research—lesson study—situated the participant and her contributions to the mathematics teachers' collaborative group.

Culturally Responsive Instruction

The socio-cultural nature of teaching and learning in a multicultural society is a call for teachers to take on a unique role of cultural broker in the classroom (Gentemann and Whitehead, 1983). As such, teachers create meaningful instructional opportunities that consider and embrace students' traditional cultural practices as part of the learning process (Gay, 2000). According to Gay, this culturally responsive pedagogy "uses ways of knowing, understanding, and representing various ethnic and cultural groups in teaching academic subjects, processes, and skills" (Gay, 2000, p. 43). Teachers who embrace culturally responsive instruction not only demonstrate increased interest in the home-based practices of the students and their families, but intentionally incorporate them in the instructional process (Hollins, 1996). They use approaches that recognize and use students' knowledge and experience gained as members of a cultural group and ensure that the social contexts of learning at school and at home are not in conflict.

In mathematics teaching, research has already found positive effects of culturally responsive instruction and professional development on student achievement and on mathematics teachers' instructional practice. Lipka and Adams (2004) studied the effects of culturally responsive mathematics curriculum supplements on students' achievement. They reported significant test score gains for all students. Their study is a signal that culturally responsive instruction in mathematics could become key for the success of ethnically and linguistically diverse students. Lipka et al. (2005a,b) investigated the effect of these practices on teachers and reported positive changes in teacher attitudes

and practices as a result of implementation of culturally responsive curriculum. These findings support the need for further inclusion of teachers in learning and teaching culturally responsive mathematics for the benefit of students and teachers.

Language Support and Beyond

Schools offer ESL services and accommodations for students with native languages different from English. These accommodations are often deemed sufficient for providing diverse students with the support they need to succeed in school; however, an array of factors adds to the complexity of schooling in a culturally different environment. Among these factors are culturally appropriate body language and presence, as well as the role and importance of rewards and personal motivation. These traditionally acceptable behaviors may differ significantly between cultures, and if ignored, could lead to misunderstandings between teachers and their ethnically and linguistically diverse students, even when the language learning process is progressing sufficiently, and influence negatively the teaching and learning processes (Ariza, 2010; Ariza et al., 2006; Zainuddin et al., 2007). Discipline problems that occur in diverse classrooms are commonly observed results of these misunderstandings.

Teachers of diverse students first need to recognize the influence of cultural norms and beliefs on learning and then work to minimize the effects of cultural discrepancies that occur in the classroom. Richard-Amato (2010) suggested using strategies and activities within the affective domain by recognizing and addressing students' attitudes, motivation, and level of anxiety. The role of ESL teachers in implementing these strategies is critical, and often it is assumed and expected that the ESL professionals are the school-wide experts on cross-cultural negotiation and understanding.

To reach every student in their classes, mathematics teachers need to acknowledge and understand that behaviors and attitudes that originate in students' native cultures may contradict teachers' own cultural expectations. As these conflicts may negatively influence the readiness and willingness of the individual student to engage with content or in certain form of learning activity in the classroom, content-area teachers need to develop knowledge and skills needed to work with ethnically and linguistically diverse students (Newman et al., 2010; Slater and Mohan, 2010). This need is critical for mathematics teachers, as erroneously, students are expected to engage with mathematical content uniformly regardless of cultural and linguistic background (Nasir et al., 2008). Researchers have recognized this as a pressing call for change—

as Newman et al. maintained, “content teachers have a need for extensive professional development in pedagogical methods and practices that have proven efficacy for ELLs [English language learners]” (p. 153). One way to build such professional development in ways that add value to teachers’ day-to-day efforts and utilize existing resources and relationships would be to team up content and ESL teachers in an ongoing collaboration.

ESL and Content Teacher Collaborations

ESL professionals today are part of most schools’ faculty. Traditionally, ESL teachers have been working in isolation of content-area teachers (Liggett, 2010), and this had a negative effect on collaborations between them, although researchers reported benefits for teachers and students when content and ESL teacher worked together. Creese (2010) observed a collaboration of geography and ESL teachers, and found that the content-area teacher struggled to recognize the input of the ESL teachers, as they were not considered content experts. Although the experience prompted both teachers to reflect on teaching practices that would benefit the linguistically diverse students, Creese found that the ESL teacher was marginalized similarly to the student population they taught. The author suggested that a preliminary recognition of the expertise of the ESL teacher was a prerequisite for a productive relationship with content teachers.

These findings were supported by Liggett (2010) who found that ESL teachers’ marginalization is in terms of both space and location of their classrooms and within the social makeup of the school. This marginalization was again reflective of attitudes toward the population they served. Liggett recommended collegial collaboration as one way for content teachers to understand the socio-cultural nature of language learning and to build appreciation and respect for the effort of ESL teachers.

Slater and Mohan (2010) explored cooperation between ESL and content teachers, and also confirmed the need for sharing knowledge and academic tasks as well as assessment strategies because they result in improved learning for students. Newman et al. (2010) described four main characteristics of such collaboration. First, content and ESL teachers needed to recognize best practices and, through cooperation, direct their colleagues toward implementing them. Second, the content teachers had to acknowledge, accept, and utilize the presence and role of cultural factors in the classroom. Third, they needed to have knowledge and understanding of language learning theories, and fourth, they

had to know and apply effective strategies on transforming their instruction and materials to be accessible to language learners.

Freeman and Crawford (2008) developed a mathematics curriculum for middle school English language learners. The authors emphasized two current barriers for the successful teaching of mathematics to non-native English speakers: first, that mathematics teachers are not trained to work with language learners, and second, that ESL teachers do not feel confident in their mathematical knowledge to teach it. Freeman and Crawford's findings are a call for finding models that recognize these key obstacles and bring mathematics and ESL teachers together in ways that best support the mathematical learning of students for whom English is not a first language.

Brooks et al. (2010) recognized that the ESL teachers have become the main resource for faculty and staff on issues related to diverse student populations, and they suggested that a professional collaboration of content and ESL teachers is "a new paradigm for teaching and serving ELL students, a paradigm in which all educators share responsibility for supporting these students" (p. 148). Identification and establishment of an effective form of mutually beneficial collaboration between ESL and content teachers is the first step toward this paradigm.

Lesson Study

Lesson study is a teacher collaboration framework that originated in Japan (Lewis, 2002; Wiburg and Brown, 2007). The typical lesson study cycle involves a group of teachers who identify a common issue in the classroom. They engage in an in-depth lesson-planning process that addresses the specific concern and often invite a knowledgeable other who contributes expertise in an area in which the participants might need extra assistance and support. The goal of the lesson-planning process is not to craft a perfect lesson, but to tailor the lesson and the ongoing teacher discussions to addressing the initial teacher concerns. One member teaches the planned lesson, while the other team members observe. A subsequent discussion often leads to lesson revisions, and, if needed, the revised lesson is taught by another team member, while the rest again observe. Another discussion and debriefing session completes one lesson study cycle.

Although lesson study has become popular in the United States only in the last decade, there are already a number of empirically supported claims that it is an effective form of mathematics teacher collaboration. Puchner and Taylor (2006) described elementary teachers' satisfaction and increased planning efficacy after participation in a

mathematics lesson study. In an examination of lesson study in mathematics with a group of elementary teachers, Fernandez (2005) focused on opportunities for teacher learning within the professional development. The overall reaction of teacher participants was that they learned much about teaching in general and about mathematics instruction in particular.

Lewis et al. (2006) reported on a six-year lesson study. The elementary grade teachers involved in this long-term project found their lesson study experience to be rich in opportunities for mentoring. Their learning was further enhanced when they invited knowledgeable others for opinion and advice. Teachers stated that lesson study was not only relevant to content-area improvement, but it was also changing the professional atmosphere at the school to one of greater collegiality and mutual support.

The variety of reported applications of lesson study suggests that the model is adjustable to the needs of teachers, schools, and students, and, as such, it was the form used by the teachers in this study. In the lesson study of this research, the study participant, an ESL teacher, worked together with three mathematics teachers.

Research Question

The guiding question for this research was: How can ESL teachers contribute to the instructional planning process for effectively teaching mathematics to culturally and linguistically diverse students? The supporting questions to this exploration were: In what ways can ESL teachers support the lesson planning and the resulting classroom practices of mathematics teachers of diverse student groups? What characteristics describe the inclusion of an ESL teacher in content planning and teaching?

Method

This exploratory case study (Yin, 1993) focused on researching the role and contributions of one ESL teacher who collaborated with three mathematics teachers engaged in a lesson study to develop instruction relevant to the diverse students in their respective pre-algebra classrooms. The study utilized data source triangulation (Denzin, 1984) that was reflective of the emergent nature of the research. Data were collected from detailed direct observations of the ESL teacher during the planning meetings with the mathematics teachers, three semi-structured interviews, and multiple informal unstructured conversations with the ESL teacher and from artifacts and documents produced or

used by the ESL teacher to explain and support the instruction-related suggestions made to the mathematics teachers.

The data analysis process was recursive, occurred throughout the study, and resulted in initial identification of emerging themes. Gradually, the accumulating data allowed for deeper pattern search that resulted in merging of initial codes and categories into more descriptive themes. After all data were collected, the selective coding process completed the data analysis through identification of the final themes and their descriptors.

Study Background and Participants

The case study participant was Gladys (a pseudonym), an ESL teacher at a middle school in the Western United States. She was a native of Australia, but she had lived and worked in Germany, Turkey, and England, where she taught English to elementary and secondary students and adults. She had also lived, studied, and worked in various parts of the United States, including Hawaii. After moving permanently to the U.S., Gladys earned a Master's degree in education, plus endorsements in ESL, mathematics, and middle-level education. As part of her studies, she completed classes on diversity in education and multicultural mathematics. Gladys had 12 years of teaching experience in the U.S.; six of them were at the middle school where the study took place. She supervised the ESL homework club and was also the academic advisor of the Multicultural Club.

The school enrolled about 1300 students in grades six through eight, with over 29% of students from diverse ethnic backgrounds. The Hispanic/Latino students formed the largest ethnic group (more than 22% of the total number of students), but a variety of ethnic groups and nationalities from around the world were also represented. Fewer than 60% of the Hispanic/Latino students met the criteria for acceptable progress in mathematics. Disaggregated data were not available for the other ethnic and linguistic groups at the school, but as a whole, the group of English language learners also did not meet the criteria.

Gladys joined a lesson study group that involved three eighth-grade pre-algebra teachers. The ethnically and linguistically diverse students in each class accounted for at least half of the total number of students. Many of these students have been or were enrolled in Gladys's ESL classes or participated in the clubs she advised. The group consecutively planned three lessons, respectively, on slope, probability, and fractions. The mathematics teachers chose and discussed the mathematical content, while Gladys brought to the teachers' attention planned lesson activities and teacher actions that appeared to

create cultural and linguistic disconnects for students who have not yet developed respective sufficient linguistic or cultural background needed to understand or complete the academic tasks and solve the mathematical problems. Each mathematics teacher then taught one of these lessons to a pre-algebra class. As a lesson study group member, Gladys fulfilled the unique role of a knowledgeable other who participated in all group activities. The teachers met once every two weeks for five months, and Gladys attended the teaching of all three lessons together with her mathematics colleagues.

Results

Gladys contributed to the lesson planning and teaching in three ways. Her input concerned, first, the parts of the planned mathematics instruction that would result in missed opportunities for learning for the ethnically diverse students because parts of the instructional materials and procedures were culturally unfamiliar to them. Second, she identified nontraditional content strategies and forms of representation that would allow students to grasp important mathematical knowledge, and third, she provided suggestions for implementation of language teaching strategies and alerted teachers of key behaviors that might signal difficulties with language instead of content. These three major themes were enhanced throughout the study by Gladys's suggestions on how mathematics teachers should maintain high expectations of all students. Her inclusion as a member of the mathematics lesson planning group revealed several characteristics of the ESL-mathematics teacher collaboration: potential for sustainability, student-centeredness, convenience, and proximity and access.

Cultural Connections

Gladys initiated discussions on topics that the mathematics teachers have not previously considered as influencing the learning of their diverse students. The planned lessons included supplementary teaching materials: worksheets, video clips, or pictures. Gladys questioned how they would help or limit students' mathematical learning or how would they enhance or limit students' learning based on cultural relevance. Early on, Gladys identified a contradiction between teachers' effort to develop activities they called "fun" and "interesting," and her observations that some of these activities are not relevant to the students. In the beginning planning steps of the first lesson on fractions, the mathematics teachers decided to use excerpts from a popular American movie, "Honey, I shrunk the kids" (Cox, 1989), to illustrate ratios and proportions. Gladys brought to the group's attention that the movie is likely

not familiar to all students, and its use would not meet the instructional purpose envisioned by the teachers. Similarly, in the lesson on slope, she asked teachers to consider that some students have not been in the U.S. long enough to relate to the examples teachers were planning to use—for example, ski slope and hiking trails. Gladys emphasized that the real-life connections should be meaningful to all students and relevant to their backgrounds for their use to result in learning. She summarized this process for the teachers: “So whatever you are using in the classroom, try and make sure that most children have things that they can bring [to the discussion]. Otherwise, they are not able to bring what they do know.”

Gladys’s suggestions also concerned the classroom management strategies used in the mathematics classrooms. She provided teachers with examples of how they influenced students’ learning and classroom participation—or allowed for students’ distraction from the learning process when not culturally appropriate. The class rules she observed in the content classrooms appeared opposite of what some children were taught at home and in schools abroad. The different, often more informal, dynamic of the teacher–student relationship in American classrooms was likely considered by some students a permission to not be engaged. As a result, students from various cultural backgrounds were often considered inattentive and failing to learn, while the reason was the lack of cultural relevance in teachers’ practice. Similarly, Gladys emphasized the importance of cultural patterns in communication with peers or large groups. She suggested that teachers assign specific student roles and provide explanation of what participation would be expected in activities. Gladys proposed that mathematics teachers observe their students in different classrooms, including their ESL classes, to gain insights on the influence of the culturally responsive classroom management on student participation and engagement with content.

Mathematics Instructional Strategies

Gladys engaged the mathematics teachers in discussions of strategies and ideas that would be understood by and relevant to all students. An immediate suggestion she made was to reconsider the use of worksheets and use manipulatives, real objects, and real situations as much as possible. She brought a variety of manipulatives she had available in her classroom—both commercial and self-made—and explained how they have been useful to support students’ understanding specifically in a multicultural classroom. She included fraction strips, blocks, and tri-dimensional cubes for the lesson on fractions. The teachers engaged in formulating and solving relevant content problems with these tools,

while they simultaneously engaged in mathematical conversations that brought to the surface language proficiency issues that affected students' understanding. Gladys suggested they used repetition and reformulation when introducing new vocabulary words as a way to build and reinforce sufficient mathematical vocabulary. She advised that students be provided with models and encouraged to negotiate meaning, clarify content, and ask content-based questions to teachers and peers.

Gladys further insisted on avoiding teaching mathematics through memorization and recall of procedures, but instead suggested creating problem-solving situations and concept development techniques. She advised against the use of a paper-and-pencil packet the mathematics teachers originally suggested for the lesson on slope and encouraged the teachers to search for a meaningful and engaging way to introduce and study slope. The team decided on a kinesthetic exploration through climbing at the school's gym. Gladys reflected on this decision:

When the teacher took them to the climbing wall, she's given them the physical experience that they can tag a memory on to, but she's now also given them a common memory, that she is now going to pull on to, so nobody would feel disadvantaged—culturally, economically, socially, because they have not ever done that [what was described in the worksheet], but they have all done this [the climbing]. So I think what they have to do is realize that this was one of the best things that [the mathematics teacher] could have done, and then we can do that with them with whatever concepts they are talking about.

In a confirmation of the value the teachers themselves found in these recommendations, they made an effort to continue applying them. The teacher who taught the lesson on slope for the group shared that after she was not able to take a class to the climbing wall, she used the school grounds for students to experience slope—and then observed:

If you can make it real life, and make [the experiences] real, and not just on the paper, it's like—oh, that building has those slopes, and then you know it is going to retain, they are going to have that background knowledge and it's going to stick.

Language of Instruction

Gladys emphasized the difference between basic social language skills and needed academic language proficiency. She encouraged

teachers to listen to students' use and understanding of academic language and encourage them to ask clarifying questions. One of the teachers shared her observations after following these suggestions. A student could not understand a mathematical task formulated using the word "similar." When the teacher inquired and explained using synonyms, the student was visibly relieved, clarified the meaning in his own words, and completed the task. As Gladys explained to the group, weaker academic vocabulary would result in guessing the meaning in context—often mathematically incorrect—or students just would not complete the task because of this language difficulty.

Gladys suggested a vocabulary review in the beginning of the lesson, with examples of word use in context followed by comparison of different uses of the same word while students share with peers and teacher. At the same time, she cautioned teachers not to single out and overemphasize vocabulary but rather advance it in unison with content learning. She was promoting a balance in mathematical and language instructional approaches that would allow students to use the English language verbally as well as in writing and drawings. In the lesson on slope, for example, students were asked to draw the walls and then write and share their own descriptive words associated with the different slopes of the walls they climbed; in the lesson on fractions, they defined fractions in their own words and then in a group developed a general definition of fractions as part of a whole. This showed the attempt to build into the lessons strategies for learning mathematical language in more accessible ways to support the understanding and participation of linguistically diverse students.

High Expectations of All Students

In this collaboration, Gladys taught, had taught, or worked with the ethnically and linguistically diverse students in their extracurricular activities. This knowledge afforded her an insider view of students' behaviors and academic performance outside of the mathematics classroom. As a result, she was able to contribute specific suggestions that were aligned with best practices in culturally responsive teaching and language teaching, yet they were made with specific observed student behaviors and abilities in mind. Gladys emphasized that all of the suggestions she made were guided by the need to put forth higher academic expectations of all students. She did not hesitate to discuss teacher actions that might be withholding learning opportunities from the diverse learners. She contrasted the behaviors of the students in their ESL classroom, where students were vocal, corrected each other, and asked teachers and peers for help, to what was observed in the

mathematics classroom, where they appeared to form a group within the class that acted quite differently. During the discussion on the lesson on fractions, she reminded the teachers that they planned for every student to verbally share their individual definition of a fraction, but in the classroom, the students were allowed to say “pass” and not have to speak if they did not want to share:

I got a chance to see my students in a mixed class with you, they are forming a layer, and it was when the microphone was passed around, - pass, pass...they are not going to be talking. When I have them in an ESL-only classroom, they can't form that lower layer.

She added that the way teachers communicate their expectations for student performance was critical, and advised, “Don't let them opt out, like the passing of the microphone...they could say pass, pass, and then those children did not get a chance to speak, because they were allowed not to.”

Gladys emphasized that teachers' sensitivity to the needs of the students was not a reason to lower expectations of their contributions and performance, as this would only put the development of both mathematical understanding and language further behind. She made a strong point for holding high expectations for the students while considering language proficiency and cultural responsiveness: “I think that we can't baby them and say—poor things, you can't do it—they can! But then make it possible for them.”

Characteristics of the Model

The effective collaboration of the teachers required mutual trust and shared responsibility for tasks and outcomes alike. In this study, teachers knew each other and were comfortable sharing, arguing, observing, critiquing, and finding solutions. Being faculty at the same school, they shared interest in the school, the students, and their own professional improvement—all needed components for a beneficial shared commitment to a professional collaboration.

Additionally, the ESL teacher was within reach for the mathematics teachers to visit, share, and ask questions even when not meeting for lesson study. The proximity of the participants was an enhancing feature of the collaboration, as it was convenient—and at the same time was moving the group forward through clarifications provided in these interim meetings. The teachers had a common frame of reference for students, ongoing events, school culture, personal student situations,

changes in student population, including newly arrived students who joined their respective classes, and any additional factors that had the potential to affect students' participation in the learning of mathematics.

Discussion

The inclusion of an ESL teacher as collaborator in the planning and teaching of mathematics lessons offers a model toward utilization of existing resources and the building of deeper professional relationships in schools. It was in unison with current recommendations for extended teacher collaboration found in the professional development literature (Glickman et al., 2007; Gordon, 2004; Guskey, 2003; Loucks-Horsley et al., 2003; Loucks-Horsley and Matsumoto, 1999; Loucks-Horsley et al. 1996). This faculty-driven form of teacher collaboration allowed for productive enhancement of existing practice that is currently sought by schools and teachers alike.

Gladys was engaged in the discipline of noticing, characterized by "increased sensitivity to notice and enhanced awareness of possibilities and choices in the moment" (Mason, 2002, p. 186) throughout her participation in the planning and teaching processes. Mason claims that this sensitivity is inseparable from being an educator, but in this collaboration, the noticing was a student-sensitive one. Gladys was sensitive to the potential effects of the actions planned by her colleagues for the diverse students in their classrooms and reflected on their suggestions through the lens of culturally responsive teaching (Gay, 2000).

This form of ESL teacher participation and input was also reflective of the nature of language learning and acculturation as ongoing processes that progress throughout the school year. The suggestions Gladys provided were relevant to the school and class populations, and she was able to speak specifically to the current levels of language development as she had direct contact with most of the linguistically diverse students on a regular basis.

A study limitation was that Gladys was not Hispanic/Latina, nor did she speak Spanish, while the majority of the diverse student population at the school was Hispanic; however, in this way Gladys represented the typical ESL educator in the nation. Additionally, in this study, Gladys was a unique participant with certain cultural and educational experiences. The outcomes of similar forms of collaboration would be influenced by the specific experiences of the ESL teachers. Another limitation here was the scope of the exploration, with only one ESL teacher and one group of mathematics teachers. Simultaneous ex-

ploration of the work of several groups will add more details and insights to this model.

Conclusion

Teachers of ESL are often the ones who have extensive knowledge and understanding of issues beyond language learning that ethnically diverse students face at school. School can utilize their expertise by creating conditions for collaboration of ESL and content teachers. The findings of this study suggest that for this collaboration to be fruitful and effective in the classroom, this person should be “Somebody who has a good rapport with the teachers already, or could develop one quickly, and who knows the makeup of the school body. So that they can be sensitive to the social, cultural, economic backgrounds of the children” (Gladys, interview three). Such reciprocal collaboration would provide much-needed sustainability of the efforts and would benefit teachers and students alike.

References

- Ariza, E. N. (2010). Not for TESOL teachers: What every classroom teacher needs to know about the linguistically, culturally, and ethnically diverse students. Boston: Allyn and Bacon.
- Ariza, E. N., Morales-Jones, C. A., Yahya, N., and Zainuddin, H. (2006). Why TESOL? Theories and issues in teaching English to speakers of other languages in K-12 classrooms. Dubuque, IA: Kendall Hunt.
- Brooks, K., Adams, S. R., and Morita-Mullaney, T. (2010). Creating inclusive learning communities for ELL students: transforming school principals' perspectives. *Theory Into Practice*, 49(2), 145–151.
- Cox, P. F. (producer) (1989). *Honey, I shrunk the kids*. [Motion Picture]. United States: Buena Vista Pictures.
- Creese, A. (2010). Content focused classrooms and learning English: How teachers collaborate. *Theory Into Practice*, 49(2), 99-105.
- Denzin, N. (1984). *The research act*. Englewood Cliffs, NJ: Prentice Hall.

Fernandez, C. (2005). Lesson study: A means for elementary teachers to develop the knowledge of mathematics needed for reform-minded teaching? *Mathematical Thinking and Learning*, 7(4), 265-289.

Freeman, B., and Crawford, L. (2008). Creating a middle-school mathematics curriculum for English language learners. *Remedial and Special Education*, 29(1), 9-19.

Gay, G. (2000). *Culturally responsive teaching: theory, research, and practice*. New York: Teachers College Press.

Gentemann, K. M., and Whitehead, T. L. (1983). The cultural broker concept in bicultural education. *The Journal of Negro Education*, 52, 118-129.

Glickman, C. D., Gordon, S. P., and Ross-Gordon, J.M. (2007). *SuperVision and instructional leadership: a developmental approach* (7th ed.). Boston, MA: Pearson

Gordon, S. P. (2004). *Professional development for school improvement: empowering learning communities*. Boston, MA: Pearson Education.

Guskey, T. R. (2003). What makes professional development effective? *Phi Delta Kappan*, 84(10), 748-750.

Hollins, E. R. (1996). *Culture in school learning: revealing the deep meaning*. Mahwah, NJ: Erlbaum.

Lewis, C. (2002). *Lesson study: a handbook of teacher-led instructional change*. Reston, VA: NCTM.

Lewis, C., Perry, R., Hurd, J., and O'Connell, P. (2006). Lesson study comes of age in North America. *Phi Delta Kappan*. December 2006, 273-281.

Liggett, T. (2010). 'A little bit marginalized': The structural marginalization of English language teachers in urban and rural public schools. *Teaching Education*, 21(3), 217-232

Lipka, J., and Adams, B. (2004). Culturally based math education as a way to improve Alaska Native students' math performance. ACCLAIM research initiative working paper No. 20.

Lipka, J., Hogan, M. P., Webster, J. P., Yanez, E., Adams, B., Clark, S., and Lacy, D. (2005a). Math in a cultural context: two case studies of a successful culturally based math project. *Anthropology and Education Quarterly*, 36(4), 367-385.

Lipka, J., Sharp, N., and Brenner, B. (2005b). The relevance of culturally based curriculum and instruction: the case of Nancy Sharp. *Journal of American Indian Education*, 44(3), 31-54.

Loucks-Horsley, S., and Matsumoto, S. (1999). Research on professional development for teachers of mathematics and science: the state of the scene. *School Science and Mathematics*, 99(5), 258-271.

Loucks-Horsley, S., Stiles, K., and Hewson, P. (1996). Principles of effective professional development for mathematics and science education: A synthesis of standards. Madison, WI: University of Wisconsin at Madison, National Institute for Science Education.

Loucks-Horsley, S., Love, N., Stiles, K. E., Mundry, S., and Hewson, P. (2003). Designing professional development for teachers of science and mathematics (2nd ed.). Thousand Oaks, CA: Corwin Press.

Mason, J. (2002). Researching your own practice: the discipline of noticing. New York: Routledge/Falmer.

Nasir, N. S., and Cobb, P. (2002). Diversity, equity, and mathematical learning. *Mathematical Thinking and Learning*, 4, (2-3), 91- 102.

Nasir, N. S., Hand, V., and Taylor, E. V. (2008). Culture and mathematics in schools: boundaries between “cultural” and “domain” knowledge in the mathematics classroom and beyond. *Review of Research in Education*, 32(1), 187-240.

Nasir, N. S., Shah, N., Gutierrez, J., Seashore, K., Louie, N., and Baldinger, E. (2011). Mathematics learning and diverse students. Retrieved from http://www7.nationalacademies.org/bose/STEM_Schools_Nasir_Paper_May2011.pdf.

National Council of Teachers of Mathematics (NCTM) (2004). The NCTM Achievement Gap Task Force: Final report. Retrieved from www.nctm.org/about/taskforce/achievement_gap.pdf.

Newman, K. L., Samimy, K., and Romstedt, K. (2010). Developing a

training program for secondary teachers of English language learners in Ohio. *Theory Into Practice*, 49(2), 152-161.

Puchner, L. D., Taylor, A. R. (2006). Lesson study, collaboration, and teacher efficacy: stories from two school-based math lesson study groups. *Teaching and Teacher Education: An International Journal of Research and Studies*, 22(7), 922-934.

Richard-Amato, P. (2010). *Making it happen: Interaction in the second language classroom, From theory to practice*, 4rd Ed. New York: Longman.

Slater, T., and Mohan, B. (2010). Cooperation between science teachers and ESL teachers: a register perspective. *Theory Into Practice*, 49(2), 91–98.

U. S. Census Bureau. (2011). U.S. Census Bureau Delivers Utah's 2010 Census Population Totals, Including First Look at Race and Hispanic Origin Data for Legislative Redistricting. Retrieved from <http://2010.census.gov/news/releases/operations/cb11-cn53.html>

Wiburg, K., and Brown, S. (2007). *Lesson study communities: increasing achievement with diverse students*. Thousand Oaks, CA: Corwin Press.

Yin, R. (1993). *Applications of case study research*. Newbury Park, CA: Sage Publishing.

Zainuddin, H., Yahya, N., Morales-Jones, C. A., and Ariza, E. (2007). *Fundamentals of teaching English to speakers of other languages in K-12 mainstream classrooms*. Dubuque, IA: Kendall Hunt.

“Quite Probable That They Will All Pass”: Teacher Training at Snow College

David Rosier

Snow College

Abstract

With the passing of the Free Public School Act in 1890, Mormon leaders in Utah feared they were being pushed out of influence, believing control over education was essentially control over the territory. Their immediate reaction was establishing teacher training academies wherein young people of their own faith could be trained, thus retaining influence while at the same time obeying the law. Two-year-old Snow Academy very quickly adapted, beginning 46 years as an effective Normal School. Snow did this with flair, embracing many new pedagogical theories, such as kindergarten and nature studies. From the beginning, Snow had a teachers' training lab, using children from the local public school. The result was the training of hundreds of young teachers. The program continued with notable success, firmly but indirectly maintaining Mormon leadership in the state. Snow discontinued its Normal program when requirements for teacher certificates were changed.

In the fall of 1901, an 18-year-old girl from Fairview, Utah, named Ellis Day began her teaching career. She had never had a day of training in teaching beyond a standard eighth-grade education except one correspondence course and two summer school courses. Nevertheless, she passed the county teachers' examination and was issued a license. She went on to a long career in education after considerably more training. She never believed she had been ready or qualified to teach that first year as an 18-year-old (Antrei, 1980, p. 319).

Sanpete County's desperate need for teachers was the reason for Ellis Day's being hired so young and so ill prepared. She was not alone. Throughout the rural counties of Utah, young people who could pass a county's teachers' examination were hired, whether or not they had had teacher training. This, of course, resulted in teachers being employed who were unprepared for the work and who did not succeed in the classroom. County officials were aware of this. Their efforts to solve the problem of unqualified teachers led them to establish an excellent teacher training program at the school now called Snow College. Convinced that education was the force by which to stabilize their society and to propel it forward, the county's leaders made a place to train teachers. Culturally and academically, the early development of Snow College's teacher training program laid a foundation that would prepare the way for five decades of teaching quality educators.

As early as 1881, the first forerunner of Snow College appeared under the name "Sanpete Educational Institute." Classes were held in Ephraim, Utah. The Institute offered a six-week summer school for people, both experienced and inexperienced, who wanted to take and pass the county teachers' examination. Sanpete County historian Albert Antrei (1980) wrote,

By the end of the century, [Sanpete Educational Institute] enjoyed an attendance of as many as one hundred teachers from Juab and Sevier Counties as well as Sanpete ... At the end of the term all prospective teachers were examined in several subjects in order to achieve certification. (p. 319)

On November 5, 1888, The Church of Jesus Christ of Latter-day Saints opened Sanpete Stake Academy in Ephraim, the school having grown in part from the success of the Sanpete Educational Institute. At first, the school was intended to serve people in Sanpete Stake, an ecclesiastical grouping of several congregations throughout Sanpete County, led by a Stake President. The school included a Preparatory Department, equivalent to fifth and sixth grades; an Intermediate Department, equivalent to seventh and eighth grades; and a Normal De-

partment, equivalent to the beginning of a teacher training program (Findlay, 1952, p. 7). Even though none of the students who enrolled were prepared for more advanced work than eighth grade, the founders of the school planned from the beginning to help resolve the desperate need for qualified teachers in rural areas. A more intense full year's study devoted to Normal training had been added by 1895. A bulletin advertising the course gives this description: "Instructions are given on the method of teaching and the principles of school management. The practice consists of the students conducting class exercises" (Barker, 1944, p. 70). "Class exercises" are now called "student teaching," a practical method of seeding quality into the school's graduates.

For people living in Utah in the 1880's, control over education was essentially control over the territory; thus, training teachers was only part of the reasoning behind establishing a school. Utah had had elementary schools from the time the first Mormon settlers entered the territory, but those schools were usually taught by unqualified teachers in schools corresponding with Church ward, or congregational, boundaries. Brigham Young, president of the Church at the time, opposed free public education, believing that an outsider should not educate a family's children (Hough, 1960, p. 117). Reports soon were made in the Eastern United States that Utah Territory was producing a generation of illiterate children. In many cases, these reports were essentially true: Attendance at these schools ran as low as 31% of school-age children (Hough, 1960, p. 117). The response, especially of Protestant churches, was to send young women who dreamed of a teaching career in the untamed west to establish schools (Enss, 2008, p. xii-xvi). These young teachers were well trained in the latest methods of education, and their schools were nearly always superior to the haphazardly staffed ones offered by the Church itself (Hough, 1960, p. 120). Such women "plied the frontiers waging war against crudeness, ignorance, crime, lawlessness and prostitution, as well as against polygamy" (Peterson, 1980, p. 297). Although the Church ended polygamy in 1890, Protestants continued believing education was the best way to defeat Mormons (Hough, 1960, p. 113, 119). As they drew students to their well-organized, modern schools, they were sure their belief was well founded.

Mormon leaders recognized this as a great threat to their way of life and knew that they must provide quality schools of their own. As early as 1866, Church apostle Orson Hyde wrote,

We would like young or middle aged men of good character ... if they are morally upright men whose orthography and pronunciation are correct—who can pass an examination in

reading, writing, geography, arithmetic and grammar ... We would prefer men of our own religious faith for teachers if they can be obtained ... A few competent female teachers could also find situations here. (Antrei, 1980, p. 318)

The strategy of placing people of their own faith in teaching positions persisted for decades and eventually led to hundreds of young Church members—males as well as many, many competent females—being trained as teachers at Snow College.

A few days before Brigham Young's death in August 1877, John Taylor was elected Territorial School Superintendent. He also succeeded Brigham Young as Church president and administered both church and schools from his office, from which he sent the most prominent Church members throughout the territory to spread his ideas (Peterson, 1980, p. 307). Taylor, like the fourth president of the Church, Wilford Woodruff, wanted schools that taught from scripture—that is, the Mormon scripture (Hough, 1960, p. 126). The school in Ephraim, whether functioning as a summer educational institute or as an academy, provided a setting in which teachers could be trained to carry out Taylor's and Woodruff's wishes.

In 1888, the year Sanpete Stake Academy was founded, the Church created the General Church Board of Education to strengthen its influence, but in 1890 such control evaporated. On February 18, 1890, the Utah Territorial Legislature passed the Free Public School Act, stating that Utah was to have schools "open for the admission, free of charge, for all children over six and under eighteen years of age, living in the district" (Hough, 1960, p. 127). The document does not define "district," but the word evolved to mean a geographic area from which children could reasonably be expected to walk to a school. The new Free Public School Act forbade Mormon doctrine or scripture to be included in schools. The Church operated only 15 academies in Utah in 1890, not nearly enough, nor nearly widely enough distributed, to reach the thousands of children it wanted to influence. Two responses with long-lasting effects came about: The Church established the system of classes that became seminaries and institutes (Quinn, 1975, p. 381); and the Church redoubled its efforts to train teachers of its own faith and to place them in the newly formed public schools. The academy in Ephraim was already positioned to aid this effort.

Sanpete Stake Academy was soon training teachers for the Church as well as for public schools. In 1890, a school bulletin announced, "teacher training [at the Academy] was to satisfy two needs: first, training the masses for participation in activities of the Church; second, supplying as many of the public-school teachers as possible" (Barker,

1944, p. 162). Despite its goal of training public school teachers, the Academy took its duty of training the entire Church membership seriously. Because many young people in central Utah had little education and even less sophistication, they were not prepared to become missionaries for the Church. Church schools rose to the occasion.

A one-year missionary course was outlined in 1900 as a course of study for "missionary students in the Church schools. No tuition was charged for this course. Subjects included in the courses were: Principles of the Gospel, Church History, Scripture Reading, and Singing" (Barker, 1944, p. 82).

This was not the only way in which Snow Academy (the name was changed in 1900) served the Church by training teachers. In the Snow Academy Catalogue for 1905–1906, this announcement was made about the then two-year-old kindergarten program:

To meet the demand for qualified teachers in the elementary grades and the Kindergarten work of the Sunday Schools, a Kindergarten Department has been added to the school ... The methods pursued will show how the children may be led to discover truths and to apply them in their own lives in a way which will result in a harmonious development of the powers of the mind. This is all done through play ... (Bulletins, 1905–1906)

The Kindergarten Department at Snow Academy was, in its time, a very modern idea. Internationally, educators had been talking and writing about this new plan for teaching young children, and American schools had accepted the idea. Utah state law required, beginning in 1903, that each school district serving a population of 2000 or more must maintain at least one Kindergarten within the district (Bulletins, Snow Academy Catalogue, 1905–1906, p. 25). From this time on, "district" came to have its contemporary meaning, an administrative office over several schools. Snow Academy had employed one of its own students to assist in training teachers to work in Kindergarten. Ellis Day, the young woman from Fairview who began her career at age 18, was at Snow both as a student and as a teaching assistant that fall. She remained there, studying and teaching, through 1906, when she graduated in the first group of students to take the Normal School specialty in Kindergarten (Bulletins, Snow Academy Catalogue, 1906–1907, p. 56). That summer she studied Kindergarten theory at the University of Chicago and also worked with Jane Addams at Hull House (Antrei, 1980, p. 319). She is listed in Snow Academy's 1906–1907 Catalogue with these details:

Ellis Day, Kindergarten. Student in the Central Utah College Summer School, Summers of 1899, 1902, 1903; Teacher in Public School, 1901–1904; Graduate of Snow Academy, 1906; Student in University of Chicago, Summer of 1906; Assistant Teacher in Kindergarten Department of Snow Academy, 1904–1906; Present position since 1906.

A major problem for school administrators during Utah's formative years was determining a teacher's qualifications. Training at Snow Academy or any other school did not necessarily mean anything to the people who appointed teachers. In fact, an educational standard had not yet been established; teachers of many kinds of educational experience were hired. Although Antrei described Ellis Day as the best-trained public school teacher in the county in her time (1980, p. 319), she and all other public school teachers had to take an annual county-administered examination before becoming certified to teach, no matter how much training they had had. This examination system had been set up before prospective teachers had much training at all, and it was very slow to go away. Each summer, county officials announced the examination in the local newspapers. In July of 1915, the Mount Pleasant Pyramid (a newspaper in another Sanpete County town) printed this news item:

An examination for the teachers of Sanpete County, who wish to secure county teachers' certificates, will be held at the Mount Pleasant Public School building on Thursday, Friday, and Saturday, July 15th, 16th, and 17th, beginning at 9 o'clock on each of the above dates.

To be eligible to enter this examination an applicant must have acquired three years of successful teaching experience, or must have completed a four year high school course in a school whose teachers meet the requirements for the state certification of high school teachers. Persons entering the examination without these qualifications will not be granted certificates.

No other examination for county certificates will be given during this year.

Following is the program of examination:

Thursday.—Morning Session: Theory and Practice of Teaching, Arithmetic.

Afternoon Session: Physiology, Orthography, Education books.

Friday.—Morning Session: Grammar, Psychology. Afternoon Session: Writing, Geography, Reading.
Saturday.—Morning Session: U. S. History, Drawing. Afternoon Session: Nature Study, History of Education.
JAMES W. ANDERSON
County Supt. of Schools
(Mount Pleasant Pyramid, July 9, 1915)

Even passing the examination did not guarantee a certificate to all who tried. In addition, those who passed the examination were required

to be found proficient in and qualified to teach the following branches of a common English education, viz: reading, writing, spelling, English grammar, geography, United States history, arithmetic, physiology and hygiene; and for a first grade certificate a candidate must pass in addition to the foregoing a satisfactory examination in civil government, physical geography, elements of natural philosophy, elementary algebra, and bookkeeping (Moffit, 1946, p. 305).

In this case, a “first grade certificate” qualified a candidate to teach the upper grades, usually grades seven and eight in a public school that went only to eighth grade.

The county teachers’ examination sounds quite daunting, but it did have a couple of escape clauses. If, for instance, a candidate did not answer the questions correctly, but the superintendent knew the applicant

to be a person of good moral character and to possess such knowledge and understanding, together with aptness to teach and govern as will enable the applicant to teach in district schools of the Territory the various branches required by law, said board of examiners shall grant to such applicant a certificate of qualification. (Moffit, 1946, p. 305)

The county board of education could revoke the license of any person who did not meet the criteria. But also, “the percentage required to pass any branch [was to] be prescribed by the board of examiners” (Moffit, 1946, p. 305). The board of examiners included the same people who needed to find enough teachers to staff their schools in the fall. So, even though the examination sounds exhausting and difficult, there

were ways, places, and times in which a candidate could be fairly sure of receiving a license to teach. At least once, in 1915, Utah teachers had a legitimate way around the exam. That was the year of the San Francisco Panama-Pacific Exposition, where Utah had a booth that its sponsors wanted Utah people to see. The Mount Pleasant Pyramid reported,

Utah teachers who visit the state's educational exhibit at the exposition this summer will be given cards certifying to their attendance at the exposition ... Presentation of the card upon return to Utah will entitle the holder to an extension of one year on a county teacher's certificate. (Pyramid May 21, 1915)

No doubt all teachers who could afford the trip traveled to California. Three days at the Exposition were surely more enjoyable than three days of examinations.

In 1917, Snow Academy became Snow Normal College, and the diplomas the school issued began to be taken more seriously by school boards. These were two-year college degrees, which could take as long as six years at Snow to earn because the college required a high school diploma to enter the program. If students did not come to Ephraim with this prerequisite met, they would find themselves taking grades nine through twelve at Snow's high school, still called Snow Academy, before entering Snow Normal College (Findlay, 1952, p. 58). Before this time, students could enter into Normal training following an eighth-grade graduation, which would mean they were mostly of high school age. "Snow Academy," stated Principal Newton Noyes, "catered to high school students because students prepared for college work were not yet available in sufficient numbers" (Barker, 1944, p. 81). This helps explain why a teacher such as Ellis Day was hired at age 18.

Although teachers could be hired while very young and with very little preparation, the training available, once an applicant received it, was extensive and rigorous. The following examination, given to Normal Department students in 1897, represents testing given over the four decades of Snow's teacher training program:

Section 1—Methods of Teaching

1. Define: (a) Education, (b) Teaching, (c) Teachers, and (d) Science. (e) Write and explain ten of the rules of catechization.

2. How have we considered the problem of Education. (b) Distinguish between Instruction and Culture, (c) Name and define the Kinds of Education.
3. Name 5 of the 10 General Principles of Education and explain each.
4. What is the Mind. (b) Define Perception, Imagination, Memory, Understanding and Reason, and tell how each can be cultivated.
5. Define Idea, Thought, Conception, Judgement. What do you understand by Inductive and Deductive Reasoning. Illustrate.
6. What is Knowledge. (b) What 2 kinds of Knowledge do you know. Define each kind. Give our Author's Classification of Knowledge and define each subdivision.
7. Explain Analytic and Synthetic, Concrete and Abstract, Inductive and Deductive, and Theoretical and Practical Instruction. Illustrate.
8. What are the 3 divisions of school life and what should be taught in each.
9. How would you teach Color to the little ones. (b) Name the Primary and Secondary Colors. (c) Explain the mixing of colors. Tell all you know about teaching Sound and Form.
10. Write an Object Lesson on the Subject of your choice, with not less than 20 leading questions and answers.
11. Give the steps to be taken in a recitation. (Sic) (Findlay, 1952, p. 111)

Snow Normal College kept current with the times, requiring acquaintance with new studies such as psychology, hygiene, and Kindergarten, adding study of these subjects to requirements for graduation. Lesser-known studies received attention, too. In 1900, Emma McVickers, the only female State School Superintendent Utah has had, criticized public schools, writing:

There is not sufficient interest in Nature Study, and in very few county schools do the children know much about the natural environment. The natural features of their district should be their geography lessons, and they should be familiar with every tree, herb, animal and insect, and intimately acquainted

with the habits of life and development. (Lubomudrov, 1981, p. 260)

It did take a while, but 19 years later, Snow added “Nature for Teachers”:

This course will supply Normals with the necessary information to inculcate into the minds of the Public School children the fundamental principles of nature work. To study from nature the flora and fauna of the locality, the geology of the surroundings, and to build up a knowledge of the economic value of animals and plants. (Bulletins, Announcement of the Snow Normal College, 1919–1920)

Another requirement was suggested by H. C. Snell, Chairman of Snow’s Normal Department in 1928. Snell insisted that “students who are being graduated with the teacher’s certificate be required in the future to: Submit a certificate in penmanship from the Palmer School or some other recognized school in penmanship” (Findlay, 1952, p. 63).

With these and other added requirements for the teaching certificate, counties finally dropped the examination and trusted the training of young teachers. However, the teaching certificate was not issued casually. Normal certificates were to be issued only to

“professional teachers who have reached the age of twenty years, have had two years successful experience in this State, and exhibit satisfactory evidence of good moral character, and upon critical examination are found to possess the requisite scholarship and culture” (Moffitt, 1946, p. 309).

The Academy had tried all along to provide teacher training for its Normal students. Principal Newton Noyes wrote,

Monday, February 13, 1899: I have succeeded today in arranging for a small class of children to attend the Academy one hour each day for the benefit of the Normal training class. This is the first step in this direction in the history of the Academy. (Noyes, n.d.)

Again in 1900, a practice class was sent over from the public school to Snow Academy (Barker, 1944, p. 81). Such practices continued, and “in 1924 the High School was moved from the college build-

ing to the Ephraim Public School and, in exchange, grades one to four ... were moved to the college building to be used as a training school for the Normal Department" (Findlay, 1952, p. 62).

Soon after the founding of Sanpete Stake Academy as a Church academy in 1888, recruitment went on even outside Sanpete Stake for all departments, not just the Normal School. By 1897, Sevier, Piute, and Wayne counties were recruited and responsive areas. "Faculty members were assigned areas to canvas for students, and were asked to speak at Church Conferences in the various Stakes to encourage students to attend the school" (Findlay, 1952, p. 78). This had been the reason the Academy's leaders had wished to change the name from Sanpete Stake Academy to something more inclusive; after all, the school was serving students from a far larger area than Sanpete Stake.

But enrollment remained a problem. In the fall of 1923, enrollment was low, so Snow, as a Church school, was permitted to require Stake Presidents from central Utah to find and send students to Ephraim at these rates: "South Sanpete, 15 students; North Sanpete, 10 students; Gunnison, 5 students; North Sevier, Sevier, South Sevier and Emery, 5 students each; Wayne and Garfield, 3 students each" (Findlay, 1952 p. 78). Of course with the Church's interest in staffing public schools with its own people, many of these prospective students were encouraged to enter the Normal program. Recruited students were promised a few immediate benefits. For several years, some stores gave students discounts, and the San Pete Valley Railway, running from Nephi to Manti, offered students reduced rates (Bulletins, Circular Snow Academy 1900–1901).

Because Newton Noyes served as education professor as well as principal at Snow, the Normal Department may well have received more attention than any other study the school offered. On Founder's Day of 1902, he wrote about a parade the school sponsored along Ephraim's Main Street: "Founder's Day, Seventeenth Anniversary. The Procession ... was as follows: the Academy Band, the Flag, the Faculty, the Business class, the Third Year Normal Class, Second Year Normal Class, and First year Normal Class" Of one graduation he wrote, "Friday, April 15, 1898: I am proud of our first graduates in the Two Year's normal course. There are fifteen in the class and it is quite probable that they will all pass." Four years later, he wrote,

Wednesday, April 23, 1902: Today the Normal Graduating Class ... planted eight choice trees on the New Academy Grounds—northeast corner. These trees were purchased by the class, and they plant them as a token of their good feelings toward their Alma Mater ... (Noyes, n. d.)

While Snow was growing from an Academy to a Normal College, one of its remarkable graduates and former faculty was changing her life too. Ellis Day, the student who also served as a teaching assistant in Snow Academy's Kindergarten Department, had married in 1909 to Leslie McLean Coombs from the Church colonies in Canada, where the two of them spent a winter. Coombs suffered from heart trouble, and doctors thought he would be better in a milder climate, so the pair moved with their first child to Colonia Juarez, one of the Church colonies in Mexico. But the change did not help for long. In February 1917, Leslie Coombs died, leaving Ellis with four small children, one of whom died in September of the same year. In October 1917, Ellis moved with her three surviving children to Fairview, where she bought a house and set about making a life for herself and her children. She began teaching at Fairview Public School in January 1918 and stayed there until her retirement in 1948. Before completing a Bachelor's degree at Brigham Young University in the summer of 1945, she supported herself and her family with credentials earned at Snow College (Durfey, 1990, n.p.).

Those credentials strongly reflect Snow's contribution to teacher training in central Utah, both culturally and academically. The school, having begun as a Church academy, trained local Mormon youth to teach in public schools, thus perpetuating the culture the school had been established to sustain. By the time Snow's title had become Snow Normal School, its programs were defining teacher qualifications, combining the scriptural strictness of its early supporters with the innovative ideas of twentieth-century educators. The school's atmosphere tempered rigorous programs with continual faculty support; as Principal Newton Noyes had written, the faculty trained them until it was "quite probable that they [would] all pass." Although the young academy struggled many times to remain open, it provided 48 years of quality teacher training.

Ellis Day Coombs was one of many teachers trained at Snow whose professional work gave the school an excellent reputation. The superintendent of Duchesne County School District once wrote the President of Snow College, asking if more Snow graduates like those he had hired the previous year could be persuaded to come to Duchesne that fall. The Carbon County Superintendent asked the same favor (Findlay, 1952, p. 122). This continued until 1936, when the state changed requirements for teachers, insisting on a four-year bachelor's degree for teaching (Findlay, 1952, p. 63), but while it lasted, from 1881 with the first Summer Institute for Teachers until 1936, Snow gave the area and the state a long line of well-trained teachers.

References

Antrei, Albert. (1980). "The Common School." In Albert C. T. Antrei and Ruth D. Scow (Eds.), *The Other 49ers*. (pp. 311–355). Salt Lake City: Western Epics.

Barker, Lincoln. (1944). *Snow College*. (Excerpt from Thesis, Maryville College, Maryville, Tennessee, 1944).

Bulletins and Catalogue of Sanpete Stake Academy, Snow Academy, Snow Normal College, and Snow College. (1894-1940). Snow College Special Collections

Durfey, Esther Coombs. (1990). "My Story." Fairview Museum, Fairview, Utah.

Enss, Chris. (2008). *Frontier Teachers: Stories of Heroic Women of the Old West*. Guilford, Connecticut: A Two Dot Book.

Findlay, Ross Partington. (1952). *Snow College: Its Founding and Development, 1888-1932*. (Thesis, Utah State Agricultural College, Logan, Utah. 1952).

Hough, C. Merrill. (1960). "Two School Systems in Conflict: 1867-1890." *Utah Historical Quarterly*, 28, 113-134.

Lubomudrov, Carol Ann. (1981). "A Woman State School Superintendent: Whatever Happened to Mrs. McVickers?" *Utah Historical Quarterly*, 49, 254-261.

Moffitt, John Clifton. (1946) *The History of Public Education in Utah*. Privately printed.

Mount Pleasant Pyramid. (1909, January through 1940, June).

Noyes, Newton B. *Diary*. (N.D.) Found in "Scribbler's Scrapbook: Items of Interest in the Development of Snow College, Part I." Snow College Special Collections.

Peterson, Charles S. (1980). "A New Community: Mormon Teachers and the Separation of Church and State in Utah's Territorial Schools." *Utah Historical Quarterly*, 48, 293-312.

Quinn, D. Michael. (1975). "Utah's Educational Innovation: LDS Religion Classes, 1890-1929." *Utah Historical Quarterly*, 43, 379-389.

References Consulted

Bradley, Martha S. (1991). "Protect the Children: Child Labor in Utah, 1880-1920." *Utah Historical Quarterly*, 59, 52-71.

Excelsior Star, The. (1899, April). Snow College newspaper. Snow College Special Collections.

Greenwood, Alma. (1888-1891). "Journal Pertaining to Sanpete Stake Academy, 1888-1891. Snow College Special Collections.

Hales, David A. (1999). "School Days and Schoolmarms." *Utah Historical Quarterly*, 67, 100-110.

Kinkead, Joyce (Ed). (1996). *A Schoolmarm All My Life*. Salt Lake City: Signature Books.

Oman, Susan Staker. (1981). "Nurturing LDS Primaries: Louie Felt and May Anderson, 1880-1940." *Utah Historical Quarterly*, 49, 262-275.

Retainer, The. (1914). Excerpts from Snow College Yearbook found in Scribbler's Scrapbook, Part III. Snow College Special Collections.

Stiles, Lindley J. et al. (1960). *Teacher Education in the United States*. New York: The Ronald Press Company.

Time for Science? A Study on the Use of Instructional Time for Teaching Science at the Elementary Level

Mary Sowder and Stan Harward

Utah Valley University

Abstract

Elementary teachers are often faced with inconsistent messages from national and local officials about the value of science instruction relative to the value of subjects emphasized on standardized tests. Despite ongoing calls from political leaders for improving student achievement in the United States, legislation mandating evaluation of schools and teachers that is based largely on results from standardized testing appears to have had some negative, if unintended, consequences for science instruction at the elementary level. This descriptive study sought to add to data from the field about the quantity and quality of elementary science teaching as sampled in the classrooms of 179 K–6 teachers across ten U.S. school districts in the Intermountain West. Quantitative evidence was gathered to show how much time was devoted to science instruction in these classrooms, especially as it compared with the amount of time devoted to instruction in subjects used to

determine adequate yearly progress (language arts and mathematics). Qualitative analysis of field observations identified the kinds of instructional strategies participant teachers used to teach science and how their lessons' content and teaching strategies were aligned with the National Science Education Standards. Results of this investigation confirmed earlier studies about the inconsistent amounts of time and resources devoted to science instruction in elementary classrooms, called into question educators' notions about the effects of devoting class time to science teaching on the amount of time available to teach tested subjects, and made a case for wider investigation into possible immediate and long-term effects of early exposure to science education on student achievement.

In assessing national progress towards implementation of their 2005 recommendations for bringing the United States back to the forefront of innovation in a global economy, the authors of *Rising Above The Gathering Storm, Revisited: Rapidly Approaching Category 5* offered four overarching recommendations, arranged in order of assigned importance. First among these was to “move the United States K–12 education system in science and mathematics to a leading position by global standards” (National Academies, 2010). Unfortunately, current multinational (e.g., Programme for International Student Assessment [PISA], 2009; Trends in International Mathematics and Science Study [TIMSS], 2007) measures of student achievement in science indicate that students in the U.S. may not be particularly well prepared to compete in science on a global level.

Political and educational leaders have called for a renewed emphasis in science instruction in U.S. public schools, but this development has been hampered to some degree by conflicting agendas toward educational reform. Increasing devotion to the use of standardized test scores in reading and mathematics to evaluate student, school, and teacher effectiveness has often led to a curricular narrowing at the elementary level to tested subjects (Au, 2009). The resulting decrease in the importance of science in the elementary curriculum seems at odds with the recommendations of the National Academies (2010). This study sought to uncover how elementary teachers in the participating schools have dealt with the dilemmas imposed by inconsistent messages about the relative value of primary science instruction from national and local officials.

The purpose of this descriptive study was to look at both the quantity and the quality of elementary science teaching as sampled in the classrooms of the teachers participating in this study. The study of

“science” at the elementary level includes learning about content information in life, earth and space, and physical sciences as well as studying the history and nature of science, science inquiry, science and technology, and science in personal and social perspectives (National Research Council [NRC], 1996, 2000). The following research questions were addressed:

- How much time do the elementary teachers devote to science instruction? How does the amount of time for science teaching compare with the amount of time devoted to instruction in other subject areas?
- What kinds of instructional strategies are used to teach science? How is elementary science instruction aligned with the National Science Standards?

Framework

This project is framed by the literature describing the state of science teaching at the elementary level and the decreasing stature of the study of science in the elementary curriculum, in light of increasing emphasis on acquiring skills in mathematics and literacy that are included on standardized tests.

The present emphasis on instruction for decontextualized reading and mathematics skills has led to narrowing of the curriculum to tested content areas, to the detriment or exclusion of non-tested subjects (Amaral et al., 2002). One nationwide survey found that 71% of school districts cut at least one subject to allow for more time to focus on math and language arts “as a direct response to the high-stakes testing mandated under NCLB [No Child Left Behind]” (Au, 2009, p. 46). This is especially true in the elementary grades, where requirements for adequate yearly progress (AYP) have placed an extra emphasis on developing students’ reading skills (Sunal and Sunal, 2007).

Science teaching has recently become an even more tangential part of the elementary curriculum because “what gets taught in a classroom is largely determined by what gets tested” (Lee and Luykx, 2006, p. 28). A report from the Center on Education Policy (2008) found that, of those schools that increased instructional time for math and reading since 2002, more than half (53%) cut time by at least 75 minutes per week in science. The curricular narrowing in response to high-stakes testing has served to deemphasize preparation for science instruction in teacher education coursework and field experiences despite persistent and increasingly alarmist calls for better-trained science teachers at all levels (National Commission on Mathematics and Science Teaching,

2000). Teacher candidates are rarely able to observe models of reform-based science teaching (as defined by the NRC, 1996, 2000) during their preservice field experiences as elementary teachers devote more and more instructional time to tested areas of the curriculum. As educational policy assigns the teaching of authentic science a low priority in the elementary curriculum (Abell and Roth, 1992), the preparation and professional development of teachers for science instruction is consequently often limited (Smith and Gess-Newsome, 2004; Smith and Southerland, 2007). Elementary teachers may enter the classroom even ill-prepared to enact effective science instruction.

In addition to concerns about the amount of time elementary teachers spend on science instruction are issues related to the manner in which that time is used. Research has shown that elementary teachers either tend to avoid teaching science (Goodrum et al., 2001; Harlan and Holroyd, 1997; Lee and Houseal, 2003; Tilgner, 1990), or they may tend to teach science in ways that are inconsistent with standards-based practice (Harlen, 1997). A consequence of these two facets of elementary science education is that there is a tendency for teachers to use teaching strategies that allow them to “maintain control of the classroom knowledge flow, but which are often not appropriate ways of engaging students in science” (Appleton, 2007, p. 497). It may be that limitations placed on time for science instruction in elementary classrooms have only exacerbated the already inconsistent practices of science teaching at the primary and intermediate levels. This study will attempt to document the amount of time devoted to science instruction at the elementary level (in comparison with other subject areas) and the nature of the instructional strategies used during that time.

Method

Participants and Context

A sample of 179 regular, full-time classroom teachers of kindergarten through sixth grade was randomly selected from 10 school districts located within one state in the Intermountain West region of the United States. Participants had varying levels of experience and educational background, and they represented a proportional sample of teachers by grade level across these districts. These teachers were observed for one entire school day of instruction by trained observers using an observation instrument. One fifth of the teachers were randomly chosen to be observed on Monday, one fifth on Tuesday, and so on during the week, so that the full week of instruction was represented.

Data Collection

Data were collected from field notes of each of the teachers as they were observed during one complete day of typical instruction by a trained observer. The observation instrument (see Appendix A) for the study was drafted, field tested, and revised prior to implementation. Once the initial draft of this form was field tested in several classrooms, revisions were made according to feedback from the testers. The feedback was also used to create instructions for training all observers, including guidelines for conducting observations as well as definitions and examples of typical classroom instructional practices and procedures observers would likely see in elementary classrooms. To protect construct validity, these definitions were used in observer training sessions and during the field observations. The observers, senior undergraduate education students from two neighboring universities, received a 90-minute training session prior to their observation work. All training sessions followed the same format—explanation of the study objectives, description of the observation instrument, practice with the instrument using video clips of classroom instruction, assessment using a video clip, and explanation of procedures to follow on the observation day. (While this instrument has not yet been validated, data from this study and other studies using this instrument might be used to provide such a validation.)

Approximately 100 trained observers individually observed one teacher for an entire school day, while approximately 17 pairs of observers observed the same teacher. Inter-rater reliability (95.5%) was established using data from the pairs of observers. Observers labeled and described each instructional activity on the observation instrument for each lesson they saw. At the end of the day, all completed observation forms were mailed to the researchers.

Data Analysis

This study used a mixed method design (Creswell and Plano Clark, 2010), because both qualitative and quantitative data were collected and analyzed. A completed observation instrument for each of the classroom observations was used by the researchers to ensure accuracy. The amount of time spent for each subject (language arts, mathematics, social studies, music, art, science, and other activities) was derived from these accounts of classroom lessons. Statistical analysis of the data was used to make comparisons of the amount of time used for instruction in each area of the curriculum. Independent samples t-tests were used to compare data from those classrooms in which science

lessons were observed and those classrooms in which no science lessons were observed.

Next, descriptions of the science lessons from the observational data were examined using a constant comparative analysis (Dey, 1993) to determine categories of teaching strategies and science content used in observed science lessons. The researchers began with an inductive, cross-case analysis of the data from descriptions of individual classroom science lessons. The process of identifying common content across the cases (Patton, 1990), or coding, allowed the researchers to group aspects of the lessons observed based on their similarities. The characteristics of each category were described to "justify the inclusion of each data bit that remained assigned to the category as well as to provide a basis for later tests of replicability" (Lincoln and Guba, 1985, p. 347). These categories then became the basis for the identification of larger themes built from an examination of the patterns in the data (Dey, 1993).

A matrix was also developed to record how these particular practices were/were not aligned with the National Science Education Standards (NSES) (NRC, 1996) for teaching, assessing, and content at the particular grade level in which they were used. A constant comparative procedure was used to build categories from the data to further illustrate the nature of the instructional strategies and methods used to teach science in these elementary classrooms.

Instructional Time for Science

Analysis of the data revealed that, among the teacher participants in this study, time devoted to teaching science was minimal in comparison with time spent on language arts lessons (reading and writing). All of the classrooms participating in the study used part of their observed instructional day learning about language arts; only a third of the teachers spent any time at all on science learning. Figure 1 illustrates the average amount of time spent for instruction across the elementary curriculum in the lessons observed for this study.

Science teaching was observed in only a third of the classrooms (59 out of 179) participating in this study. The amount of time devoted to teaching different subjects varied somewhat from grade level to grade level, but the percentage of classrooms in which at least one science lesson was observed was greatest in the intermediate grades (4–6) (Table 1).

Of those teachers who spent time teaching science, however, the amount of time devoted to science instruction did not vary significantly.

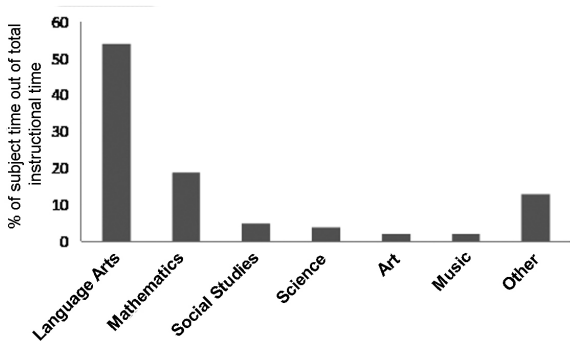


Figure 1. The percent of instructional time observed for each subject out of the total amount of instructional time available.

The average number of minutes spent teaching science, for those teachers who devoted time to science instruction, ranged from 30 minutes to 43 minutes (Figure 2). There was no clear pattern of relationships between the number of minutes in a lesson and the grade level of the student. Grade six had the highest number and kindergarten classes had the lowest number of minutes; however this may have been due to the lower number of instructional minutes available in the kindergarten school day.

Table 1. Percent of Teachers Teaching Science at Each Grade Level							
Grade level	K	1	2	3	4	5	6
Percent teaching science	22	16	22	19	34	46	72

Note: n=59

The data also revealed an interesting finding related to time spent on science instruction and the number of minutes spent teaching mathematics. For teachers teaching science, the average number of minutes spent teaching mathematics was higher than for teachers who did not teach science (Table 2). An independent samples t-test of the data for each group of classrooms (science, no science) showed a significant difference in the amount of time spent on teaching mathematics.

A common concern expressed by elementary teachers is that taking time to teach science causes a reduction in instructional time avail

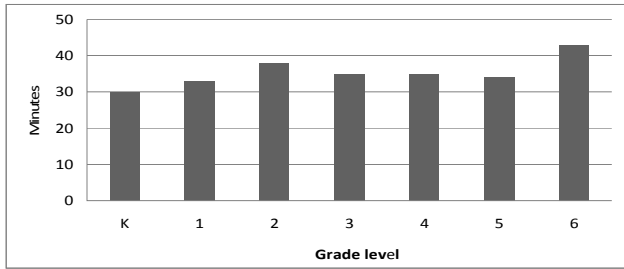


Figure 2. The average number of minutes spent on science lessons for those classrooms in which science was taught.

able for teaching those subjects (especially reading and mathematics) that are measured by standardized tests to determine acceptable rates of AYP (Spillane et al., 2001). While the data above seem to contradict this assumption for mathematics, further analysis may initially appear to support this concern when it comes to language arts. Even though time allotted to teaching language arts formed by far the largest part of instruction at every grade level, statistical analysis suggests that, at the surface level, those teachers teaching science generally spent less time

Table 2. Difference in Time Spent on Mathematics Instruction		
Group	Minutes for math instruction, with science instruction	Minutes for math instruction, with no science instruction
Mean	61.61	51.33
Standard deviation	29.44	29.4
Standard error of the mean	3.83	2.68
No. of participants	59	120

Note: two-tailed P value=0.0293; degrees of freedom=177; t-value=2.2

teaching language arts in terms of formal lessons focused only on reading and writing skills (Table 3). An independent samples t-test comparing the two groups indicated that the difference in time spent on teaching language arts was very significant.

Table 3. Difference in Time Spent on Language Arts

Group	Minutes for language instruction, with science instruction	Minutes for language instruction, with no science instruction
Mean	142.63	167.93
Standard deviation	48.78	58.34
Standard error of the mean	6.3	5.33
Number of participants	59	120

Note: two-tailed P value = 0.0046; degrees of freedom=177; t-value=2.9

What this analysis does not take into account is the nature of the science lessons observed for this study. All of the science lessons addressed, more or less, objectives outlined by the state core curriculum; however many of these lessons were, in whole or in part, more closely related to instruction for reading and writing about science topics than to instruction in developing conceptual understandings of science content and processes. Even those lessons teaching science as inquiry required students to communicate their understandings in spoken or written form. The following section will present a qualitative analysis of the nature of the science lessons observed.

Teaching Elementary Science

This analysis is, by necessity, limited to data from the 59 (out of 179) classrooms in which science lessons were observed for this study. The qualitative data supplied by the observers' field notes of science lessons were reviewed by the researchers and categorized in two ways. A constant comparative method was used to code the data according to the kind of activities and instructional methods used during the lesson. The data were also organized according to how they addressed elements of the NSES (NRC, 1996).

Categories of instructional methods. Descriptions of classroom science activities collected were used to develop a representation of the kinds of common instructional practices observed across science lessons. The lessons analyzed for these categories did not confine them-

selves to only one instructional method included in the matrix analysis (Table 4). A single lesson may have used a number of instructional practices, all of which were counted in the analysis. For example, students investigating the properties of water might also produce written descriptions of their observations; students listening to a presentation of information might also be required to respond in writing to questions on a worksheet or test.

The first standard listed for science education (NSES, 1996) is Science as Inquiry, and more than a third of the science lessons in this study included activities that at least partially addressed the skills and processes outlined in the standard. This standard was often addressed only in part, as an activity included within a larger lesson that was not necessarily based on the inquiry model of instruction. When examined in the light of the total number of lessons observed in all 179 classrooms, the process of science inquiry was only observed in 2% of the lessons.

Science lessons and the NSES. All but three of the lessons observed used a direct instruction model for teaching science; the other three lessons were taught with a guided inquiry approach. Even though direct instruction was used as the primary method of delivering information, many of these lessons also incorporated activities that addressed at least some of the inquiry skills listed in the NSES (NRC, 1996). For example, one of the lessons that was primarily focused on a discussion of science fair projects did so in a way that emphasized this forum as similar to the way “scientists make the results of their investigations public in ways that enable others to repeat the investigations” (NRC, 1996, p.123). Another direct instruction lesson about the solar system ended with an opportunity for students to do follow-up work on group investigations into students’ questions, “seeking information ... from their own observations and investigations” (NRC, 1996, p. 122). Figure 3 presents data on how the observed science lessons addressed the six content standards from the national standards.

The percent of science lessons teaching any particular content area was affected by the differences in the number of teachers teaching science at each grade level. The highest number of teachers who devoted some time to science instruction taught intermediate grades, especially at the sixth-grade level. Because one of the major areas of study outlined in the state’s requirements for science instruction at this level was earth and space science, there was a higher percentage of lessons addressing this content.

Table 4. Analysis of Approaches to Science Instruction		
Category	Description	Lessons using method
Teacher-centered instruction	Relied heavily on teachers reading, talking, or demonstrating to the whole class. Included lessons that involved shared reading (the teacher reading to the students), depended on lecture or direct instruction for delivery of information, or may have incorporated teacher-led demonstrations of scientific phenomena.	94%
Students speaking, reading, or writing about science (language arts)	Showed students reading paper or online informational texts, listening and taking notes of videos, talking with one another or presenting information to the class, learning new vocabulary, composing written responses or summaries, or filling out worksheets or organizers.	88%
Student investigations/inquiry	Included activities in which students were actively engaged with materials first-hand to observe, measure, and record data, create hypotheses, or identify evidence. Required students not only to manipulate materials, but to thoughtfully reflect on and interpret their observations. (Activities requiring active participation without reflection were included in “other instructional activities” described below.)	37%
Content tests	Included, or were limited to, completion of a formal, summative evaluation of learning about science content.	15%
Other instructional activities	Encompasses other, content-weak activities used during science lessons, including students eating, playing games, and participating in craft activities loosely related to a science topic.	12%
Discussions of science fair	In some classrooms, instructional time for science learning was devoted to, or included, discussions of science fair entries.	8%

Discussion

Data collected from the classroom observations illustrated practices for teaching elementary science in these classrooms that are both disconcerting and encouraging. Fewer than one third of the classrooms observed devoted any instructional time to science learning. Teachers in the classrooms observed for this study spent almost 14 times the amount of time on teaching language as they did on teaching science, even without figuring in the minutes during science lessons that were really focused on language arts activities. While teaching young stu-

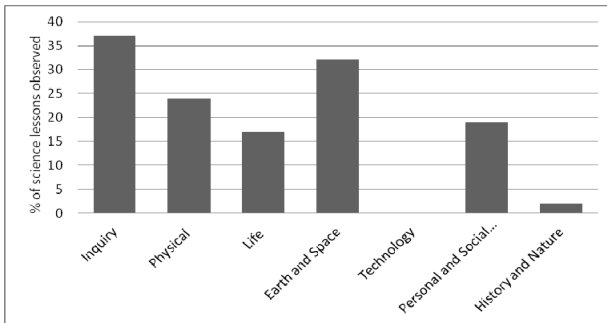


Figure 3. Alignment of science lesson content with the six content standards from the NSES (NRC, 1996).

dents how to read, write, and speak is central to the elementary curriculum, an overemphasis on the use of instructional time for these subjects appears to have led to a corresponding under emphasis on other areas, including science, that do not form part of high-stakes testing used to determine schools' AYP. These findings align with those from earlier studies on curricular narrowing as a consequence of a reliance on high-stakes testing as the central measure of student achievement (Au, 2009; Center on Educational Policy, 2008).

This study also provides some evidence that elementary educators and administrators may have some naïve ideas about the effect of science teaching on the amount of time available for teaching language arts and mathematics. Analysis of the data in this study appears to indicate that devoting time to science teaching does not lead to a decrease in the amount of time devoted to teaching mathematics and that most science lessons also incorporate the teaching of reading and writing within the content area. While the finding that those teachers who taught science also increased time for mathematics instruction may

have been related to those teachers' level of confidence or comfort teaching math (and vice versa), it may also be that learning about science content requires greater mathematical understanding on the part of the students. Furthermore, because science lessons can create authentic contexts for the application of skills taught in more decontextualized language arts and mathematics lessons, they may serve to enhance student learning in these subjects (Brown et al., 1989; Lave, 1988; Lave and Wenger, 1990; Suchman, 1988).

Nature of Elementary Science Instruction

Data analyzed in this investigation also revealed that while most of the science lessons observed relied heavily on teacher-centered learning activities, a little over a third of them also incorporated opportunities for students to participate in inquiry learning. Even though the percentage of lessons using science inquiry was very small in comparison to the total number of lessons observed overall (about 2%), the data indicated that there has been some movement toward more standards-based instructional practices (as described in the NSES) in elementary science teaching.

Of greater concern may be the data showing that none of the science lessons observed in this study addressed the NSES, Standard E, for technology and science. None of the lessons observed posed challenges related to technological design or asked students to engineer solutions relating to human problems or needs (NRC, 1996). This omission may have been related to the nature of the standards for science teaching created for the state in which this study took place. The state standards for technology and science at the elementary level were limited to teaching about technology only as it was related to learning about tools used for scientific discoveries and data collection in other science content areas. National efforts to engage greater numbers of students and teachers in education for science, technology, engineering, and mathematics appear to have had little influence on the focus of local teaching standards and, consequently, classroom instruction at the elementary level.

Effects of Elementary Science Education

The apparent lack of deliberate and consistent teaching of science content and the inquiry process at the elementary level as evidenced in the results of this study gives rise to important questions about the effect this may have on students' future learning in science and across the curriculum (Fleischman et al., 2010). In the 2009 PISA results, American students ranked 21st out of 33 in science literacy among students

from developed countries (National Center for Education Statistics, 2006). It is important to note here relevant limitations of the PISA study that may have impacted these rankings. The results of this international assessment should be considered in light of cultural factors in different countries that may influence educational policy and practices, including policies affecting the inclusivity and exclusivity of educational systems. These results are also influenced by socioeconomic factors within and among the tested countries.

The 2009 National Assessment of Educational Progress (NAEP) reported that only 34% of the fourth graders, 30% of eighth graders, and just 21% of twelfth graders were rated “proficient” or better in science. One percent of American fourth and twelfth graders earned an “advanced” score, as did 2% of the eighth graders (NAEP, 2009). Taken together with findings from this study about the amount of instructional time and the nature of science education in the elementary classroom, these scores may indicate that the role of science education in the elementary curriculum should be revisited and revised. Just as it would be unreasonable to expect students to suddenly become proficient at reading, writing, or mathematics in middle or high school without some preparation in those subjects in the lower grades, so it appears, at least from the data from national and international testing programs (NAEP, 2009; PISA, 2009; TIMSS, 2007), that a lack of preparation in science learning at the elementary level may be affecting U.S. students’ future achievement.

This achievement may also be suffering from the narrowing of the science curriculum to accommodate minimal elementary science education standards among the various states. Amid popular complaints about the decreasing amount of innovation and invention from students educated in the U.S., the data showed that virtually no efforts were being made during the science lessons observed to address the NSES (NRC, 1996) for technology. This standard, absent from the elementary science curriculum for the state in which this study was situated, asks students to develop abilities of technological design: identifying a simple problem, proposing and implementing solutions, and evaluating and communicating about products, designs, and solutions—skills critical to success in a global, technology-dependent environment.

Future Directions for Research

Science education is uniquely qualified to offer opportunities for even very young students to develop skills in problem-solving and technological design. Teaching elementary students about how to use the processes of scientific investigation can provide a platform for the

kind of inquiry learning (NRC, 1996) that might help students begin to understand how to evaluate and use information to solve problems. It might also, as did some of the lessons observed for this study, give students opportunities to apply discrete bits of knowledge from other subject areas acquired in less authentic contexts.

Whether this will improve student scores on the standardized tests used for evaluating systems of education, schools, and teachers is an important consideration for future research. The results of this study indicate that a more balanced approach to enacting the elementary curriculum might be an important first step in determining the possible benefits of devoting more time to instruction in science at the elementary level, but there seems to be little political appetite for taking the risk to investigate this approach. There is virtually no research available that provides data about student achievement and time spent on elementary science instruction among countries, states, or schools that are participating in comparative studies of student achievement in science at the elementary level. An investigation into the effects of limited instructional time on science achievement, even at a local or national level, may need to await a shift in educational sensibilities. The current disproportionate emphasis in the elementary classroom on teaching tested content areas has served to make science instruction a marginalized area of the curriculum.

References

- Abell, S. K., and Roth, M. (1992). Constraints to teaching elementary science: A case study of a science enthusiast student teacher. *Science Education*, 76, 581-595.
- Amaral, O., Garrison, L. and Klentschy, M. (2002). Helping English learners increase achievement through inquiry-based science instruction. *Bilingual Research Journal*, 26(2), 213-239.
- Appleton, K. (2007). Elementary science teaching. In Abell, S. & Lederman, S. (eds.), *Handbook of research in science teaching* (pp. 493-535). Mahwah, NJ: Lawrence Erlbaum Associates.
- Au, W. (2009). Social studies, social justice: W(h)ither the social studies in high-stakes testing? *Teacher Education Quarterly*, 36, 43-58.
- Brown, J.S., Collins, A. and Duguid, S. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32-42.

Center on Education Policy. (February, 2008). Instructional time in elementary schools: A closer look at changes for specific subjects. *A report in the series, From the Capital to the Classroom: Year 5 of No Child Left Behind Act*. Washington, DC: Center on Education Policy.

Creswell, J. W., and Plano Clark, V. L. (2010). *Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA: Sage.

Dey, I. (1993). Creating categories. *Qualitative data analysis* (pp. 94-112). London: Routledge.

Fleischman, H.L., Hopstock, P.J., Pelczar, M.P., and Shelley, B.E. (2010). Highlights from PISA 2009: performance of U.S. 15-year-old students in reading, mathematics, and science literacy in an international context (NCES 2011-004). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Goodrum, D., Hackling, M., and Rennie, L. (2001). *The status and quality of teaching and learning of science in Australian schools*. Canberra, ACT: Commonwealth of Australia.

Harlen, W. (1997). Primary teachers' understanding in science and its impact in the classroom. *Research in Science Teaching*, 36(3), 323-337.

Harlen, W. and Holdroyd, C. (1997). Primary teachers' understanding of concepts of science: Impact on confidence and teaching. *International Journal of Science Education*, 19, 93-105.

Lave, J. (1988). *Cognition in practice: Mind, mathematics, and culture in everyday life*. Cambridge, UK: Cambridge University Press.

Lave, J. and Wenger, E. (1990). *Situated learning: legitimate peripheral participation*. Cambridge University Press.

Lee, C.A. and Houseal, A. (2003). Self-efficacy, standards, and benchmarks as factors in teaching elementary school science. *Journal of Elementary Science Education*, 15(1), 37-55.

Lee, O., and Luykx, A. (2006). *Science education and student diversity*. New York: Cambridge University Press.

Lincoln, Y. S., and Guba, E. G. (1985). *Naturalistic inquiry*. Newbury

Park, CA: Sage.

National Academies. (2010). *Rising above the gathering storm, revisited: Rapidly approaching category 5*. Washington, DC: National Academy Press.

National Assessment of Educational Progress (NAEP). (2009). *Science: summary of major findings*. Washington, DC: US Department of Education. Available at http://nationsreportcard.gov/science_2009/summary.asp. Accessed March 15, 2011.

National Center for Education Statistics (NCES). (2006). *Highlights from PISA 2006: Performance of U.S. 15-year-old students in science and mathematics literacy in an international context*. Washington, DC: US Department of Education. Available at <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2008016>. Accessed March 15, 2011.

National Center for Education Statistics (NCES). (2007). *Trends in international mathematics and science study (TIMSS)*. Available at <http://nces.ed.gov/timss/index.asp>. Accessed March 15, 2011.

National Commission on Mathematics and Science Teaching. (2000). *Before it's too late: A report to the nation from the National Commission on Mathematics and Science Teaching for the 21st Century*. Washington, DC: U.S. Department of Education. Accessed August 4, 2008.

National Research Council (NRC). (1996). *National science education standards*. Washington, DC: National Academy Press.

National Research Council (NRC). (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. Washington, DC: National Academy Press.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage Publications.

Programme for International Student Assessment (PISA). (2010). *PISA 2009 results: What students know and can do: student performance in reading, mathematics and science (volume I)*. Camberwell, Victoria: Australian Council for Educational Research and Organisation for Economic Co-operation and Development.

Smith, L. and Gess-Newsome, J. (2004). Elementary science methods courses and the National Science Education Standards: Are we adequately preparing teachers? *Journal of Science Teacher Education*, 15(2): 91-110.

Smith, L. and Southerland, S. (2007). Reforming practice or modifying reforms?: Elementary teachers' response to the tools of reform. *Journal of Research in Science Teaching*, 44(3), 396-423.

Spillane, J., Diamond, J., Walker, L., Halverson, R., and Jita, L. (2001). Urban school leadership for elementary science instruction: Identifying and activating resources in an undervalued school subject. *Journal of Research in Science Teaching* 38(8), 918-940.

Suchman, L. (1988). *Plans and Situated Actions: The Problem of Human/Machine Communication*. Cambridge, UK: Cambridge University Press.

Sunal, C. S., and Sunal, D. W. (2007). Reports from the field: Elementary teacher candidates describe the teaching of social studies. *International Journal of Social Education*, 22, 29-48.

Tilgner, P.J. (1990). Avoiding science in the elementary school. *Science Education*, 74, 421-431.

Appendix A

Classroom Instruction Observation Form

Teacher #: _____
 Observer #: _____ Date: _____

Activity Label	Start time	Stop time	# of students
Description			

Side-force Amplification on an Aerodynamically Thrust Vectored Aerospike Nozzle

Shannon D. Eilers, Matthew D. Wilson, Stephen A. Whitmore, and Zachary W. Peterson

Utah State University

Abstract

Results from numerical and cold-flow experimental investigations of aerodynamic thrust vectoring on a small-scale aerospike thruster are presented. Thrust vectoring was created by the injection of a secondary fluid into the primary flow field normal to the nozzle axis. The experimental aerospike nozzle was truncated at 57% of its full theoretical length. Data derived from cold-flow thrust vectoring tests with carbon dioxide as the working fluid are presented. Injection points near the end of the truncated spike produced the highest force amplification factors. Explanations are given for this phenomenon. For secondary injection near the end of the aerospike, side-force amplification factors up to approximately 1.4 and side-force specific impulses up to approximately 55 s with main flow specific impulses clustering around 38 s were demonstrated. These forces crisply reproduce input pulses with a high degree of fidelity. The side-force levels are approximately 2.7% of the total thrust level at maximum effectiveness. Higher side forces on

the order of 4.7% of axial thrust were also achieved at reduced efficiency. The side-force amplification factors were independent of operating nozzle pressure ratio for the range of chamber pressures used in this test series.

I. Introduction

I.A. Aerospike Nozzle History

The first major round of testing was completed on aerospike nozzles in the 1950s and 1960s when truncated plug nozzles were under consideration for the Saturn V upper stages,¹ and then later the Space Shuttle's main engine.²⁻⁴ During this period, Rocketdyne conducted extensive research into both aerospike performance and liquid injection thrust vectoring capability.^{5,6} As a result of this test series, Rocketdyne concluded that Aerospike nozzles had less or equal thrust vectoring capability than bell nozzle counterparts; however, their tests were limited to liquid injection as they did not perform cold flow thrust vectoring tests, and hot gas injection hardware was not yet available. After a conventional bell nozzle was chosen for the Space Shuttle Main Engine, work on aerospike nozzles decreased substantially until the 1990s, when work began on the X-33 single stage to orbit concept.⁷ To further this effort, additional work and testing was performed by Rocketdyne for Lockheed during the development of the RS 2200 linear aerospike.⁸⁻¹⁰

After the development of the X-33 and the Venture Star was canceled, work on aerospike nozzles once again became more sporadic. In America, National Aeronautic and Space Administration (NASA) Langley has worked on further parametric modeling and optimization of aerospike nozzles.¹¹ Analytical research including developing computational algorithms to evaluate thrust vector control for aerospike nozzles was performed at the University of Huntsville,¹² and differential throttling research has been completed under the direction of Marsh Space Flight Center.¹³ Engineers at NASA Dryden and Air Force Research Laboratory designed and flew an aerospike nozzle on a high-power rocket.¹⁴ Attempts have been made to develop annular aerospike nozzles for hybrid rockets at Arizona State University¹⁵ and the University of Washington,¹⁶ although there were notable challenges because erosion of the nozzle support structure occurred in the former and nozzle ablation rates were not presented in the latter. California Polytechnic University has also investigated coupling an aerospike nozzle with a hybrid rocket motor. Their efforts centered on active cooling techniques.^{17, 18} California State University, Long Beach in association

with the Garvey Spacecraft Corporation has also completed extensive testing of liquid, clustered aerospike engines, which have culminated in the launch of several sounding rockets.¹⁹⁻²³

Outside of the United States, aerospike nozzles have enjoyed a large amount of attention in recent decades. The European Space Agency has investigated the relative effectiveness of various aerospike thrust vector control techniques.^{24, 25} In the mid 1990s, The Technical University of Munich performed analytical research on performance aspects of aerospike nozzles, including performance losses due to nozzle clustering.^{26, 27} Some research has also been completed at the German Aerospace Center, mostly concerning effects of aerospike cluster configurations.²⁸⁻³⁰ A substantial amount of work has also been completed in Italy on performance validation, flight behavior, and motor cluster performance for aerospike nozzles.³¹⁻³⁵

A great deal of analytical work has been performed at several Universities in Japan on aerospike performance, slipstream effects, slipstream effect mitigation, and base bleed injection.³⁶⁻⁴⁴ Experimental work has also been completed in Japan in an effort to investigate the flow field of clustered linear aerospike nozzles.⁴⁵ Research leading to conceptual level design work for an aerospike nozzle to support single-stage to orbit vehicle design has been undertaken recently at the Japan Aerospace Exploration Agency.⁴⁶⁻⁴⁸ Beijing University in China has performed analysis as well as cold flow tests on aerospike nozzles investigating nozzle performance, base bleed effects, and thrust vectoring,⁴⁹⁻⁵³ mostly in regard to linear aerospike engines.

The National Aerospace Laboratories in Bangalore, India, have investigated the acoustics of aerospike nozzles⁵⁴ and performance characteristics of conical aerospike nozzle contours.⁵⁵ Some analytical work has also been completed in Russia on optimal aerospike contours.⁵⁶ The Aerospace Research Institute in Iran has also completed some work on base bleed performance.⁵⁷

I.B. Potential Aerospike Nozzle Space Applications

While aerospike nozzles have long been known for their altitude compensation ability during endo-atmospheric flight,⁵⁸ they also present significant potential advantages for purely in-space applications. Aerospike nozzles can be both more efficient and significantly smaller than conventional high-expansion-ratio bell nozzles. Given a fixed vehicle base area, an aerospike nozzle can present higher area expansion ratio than a bell nozzle, providing better performance in a space environment or near vacuum environment like Mars. The increased specific impulse (Isp) due to a higher possible expansion ratio using an

aerospike nozzle translates to a 8–9% decrease in the propellant mass and total system weight for space and near-space applications.¹⁴

Additionally, one of the often-overlooked properties of the aerospike nozzle is the ability to achieve thrust vectoring aerodynamically without active mechanical nozzle gimbals. This offers a significant potential for reduced system complexity and weight. In contrast to secondary injection in a bell nozzle, thrust vectoring performed by secondary fluid injection on an aerospike nozzle could also be used for attitude control independent of main thruster operation, albeit at reduced effectiveness. This makes fluidic thrust vectoring on aerospike nozzles a potential replacement for both gas attitude control thrusters and main engine thrust vector control, yielding even greater advantages in system mass and simplicity.

Figure 1 compares two aerospike-based nozzle designs with their conventional counterparts with the same effective expansion ratios. Figure 1a compares the original Saturn V first stage F-1A engine to its proposed replacement J-2T-250K aerospike engine (featuring a truncated plug nozzle). It should be noted that both of these engines are optimized for earth atmospheric conditions. At Mars and vacuum conditions, the size difference is greater. This size difference is illustrated by Figure 1b, where the proposed 12.1-kNt Altair Lander Engine is compared with its aerospike equivalent. In both examples, the size differences are pronounced.

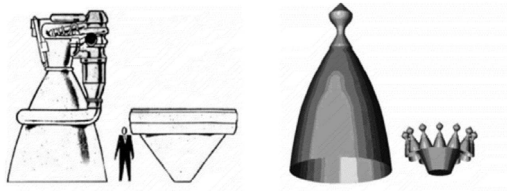


Figure 1. Aerospike compared to conventional nozzles. a) Comparison of F-1 engine to proposed J-2T-250K aerospike. b) Comparison of 12.1-kNt lunar ascent engine for Altair Lander with proposed aerospike design.

Despite these potential benefits over conventional conical or bell-nozzle designs, because of a perceived low technology readiness level the aerospike rocket configuration has never been deployed on an operational space vehicle. One of the major reasons for this perception is the lack of high-quality ground and flight test data and its correlation with analytical flow predictions. This dearth of data is especially true

with regard to off-nominal design performance, thrust vectoring, and thruster-out scenarios for clustered aerospike configurations.

II. Cold Flow Testing

The primary objective of cold flow aerospike testing was to examine the viability of fluidic thrust vectoring by gas injection on a truncated annular aerospike nozzle in near-optimally expanded conditions. Although the final aerospike nozzle was slightly over-expanded, it was designed such that the near-surface flow field was nearly identical to flow fields that would be experienced by an under-expanded or optimally expanded nozzle. In this way, the thrust vectoring research centered on thrust vectoring that would be primarily applicable to high altitude or in-space conditions.

Research emphasized examination of the effect of injection location on thrust vectoring effectiveness as well as side-force fidelity and dependence on the nozzle pressure ratio.

II.A. Cold Flow Experimental Setup

All aerospike static tests were performed in the Engineering Technology Department's Jet Engine test cell on the Utah State University (USU) campus. For static thrust tests, commercially available test stands were examined and found to be excessively expensive and have structural support mechanisms that were unsuitable for mounting the aerospike prototype. Consequently, a custom-made, portable, test stand was designed and built to support the needs of the aerospike project.

The test stand features a six-degree-of-freedom load balance with Omega LCCD type S load cells configured as shown in Figure 2. Three 100 lbf-range axial and three 25 lbf-range lateral load cells are arranged such that six-degree-of-freedom force and moment measurements can be resolved. The thrust stand coordinate system, pictured in Figure 2, is defined with x-axis vertically upward along the axial centerline of the nozzle. The thrust stand is designed so that the nozzle exhaust plume exits vertically, and the thrust acts downward onto the test cart. The test stand was calibrated in situ with a simultaneously multiaxial calibration method. The total resultant uncertainty (to 95% confidence) for forces using this calibration method was statistically determined as approximately 0.25 newtons for side forces and 1.75 newtons for axial loading.

For ease of storage, carbon dioxide was chosen for a working fluid. Figure 3 presents a schematic diagram of the associated cold-gas feed system. Saturated liquid carbon dioxide is stored in standard K-sized storage tanks, with each tank having a storage capacity of approximately 25 kg. Multiple tanks were manifolded to assure that the

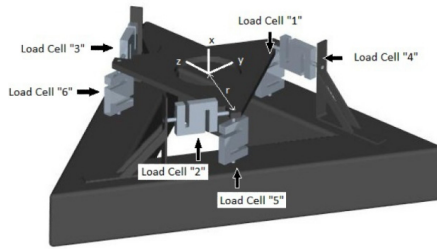


Figure 2. Six-degree-of-freedom test stand.

required mass flow levels and run times can be achieved. Flow out of the tanks is controlled via a pneumatic ball valve. The pneumatic valve actuator is controlled with a 12-volt direct current (DC) solenoid valve. Beyond the ball valve, carbon dioxide flows through a manually set needle valve that drops the pressure from the saturation pressure of carbon dioxide, 4825–5515 kPa (700–800 psia) at room temperature, to approximately 1035 kPa (150 psi). Carbon dioxide then flows into a water-bath heat exchanger, which raised the temperature of the expanded carbon dioxide by approximately 25°C. The pressure downstream of the needle valve is controlled using a back-flow pressure regulator and a primary regulator in parallel. This configuration maintains approximately 1034 kPa (150 psi) upstream of the primary regulator. The primary flow regulator then further drops the feed pressure to approximately 690 kPa (100 psi) at the plenum inlet.

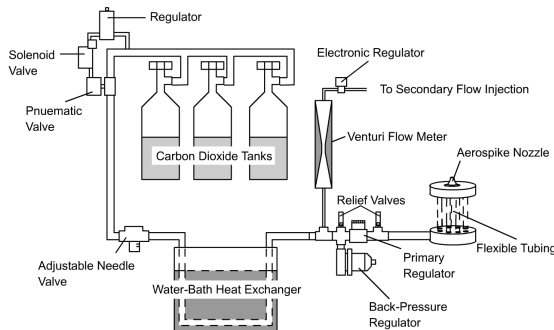


Figure 3. Aerospike propellant feed system.

At full pressure, the primary regulator is set to allow approximately one kilogram per second mass flow through the aerospike nozzle throat. The back-pressure regulator will vent approximately half that flow rate at start up. As the tanks evacuate and the overall system pressure drops, flow through the back-pressure regulator diminishes to zero. An additional electronic regulator in parallel with the main flow regulator controls the upstream pressure of the secondary (thrust vectoring and base-bleed) flow injection ports.

Type K thermocouples and Omega PX400 series pressure transducers are used to monitor temperatures and pressures throughout the flow system. A custom manufactured venturi flow meter, also using Omega PX400 pressure transducers to measure the pressure differential, is situated upstream of the electronic regulator. Although a differential pressure transducer was not used, the pressure transducer voltage bias is removed at full operating pressure when the secondary flow injection is turned fully off which results in a highly accurate differential pressure measurement. The venturi was calibrated in situ using high-flow coefficient sonic orifices. In this manner, the flow coefficient for the venturi was calculated to be 0.980.

II.B. Test Article Description

The aerospike used for cold flow testing was sized such that it was slightly overexpanded for operating conditions at the test altitude in Logan, Utah (Table 1). As the test conditions were created to approximate space conditions, the aerospike was designed using a method of characteristics code such that compression waves generated by over

Table 1. Cold-Flow Aerospike Parameters	
Aerospike Parameter	Value
Plug diameter	3.2 cm
Outer throat diameter	3.86 cm
Truncated length	2.54 cm
Full isentropic spike length	4.31 cm
Truncation ratio	57%
Throat diameter	0.29 cm
Operating stagnation pressure	775 kPa
Nozzle expansion ratio	2.47
Plenum exit throat area	4.73 cm ³
Secondary injection port diameter	0.3175 cm
Design altitude	4206 m
MSL design thrust	454 N

expansion would not intersect the end of the truncated spike at full chamber pressure. This resulted in an aerospike pressure distribution roughly independent of atmospheric pressure except for the base area and the very end of the spike length. This effect was confirmed through the use of computational fluid dynamics. The aerospike and plenum are shown in Figure 4.

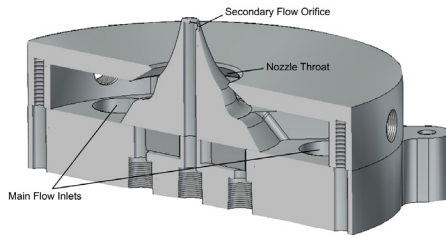


Figure 4. Cold flow aerospike test configuration.

II.C. Cold Flow Test Results

Aerospike configurations with injection ports at 20%, 80%, and 90% of truncated length were tested with secondary mass flow rates between 0.005 kg/s and 0.016 kg/s. These flow rates correspond to secondary flow inlet pressures between approximately 400 and 800 kPa.

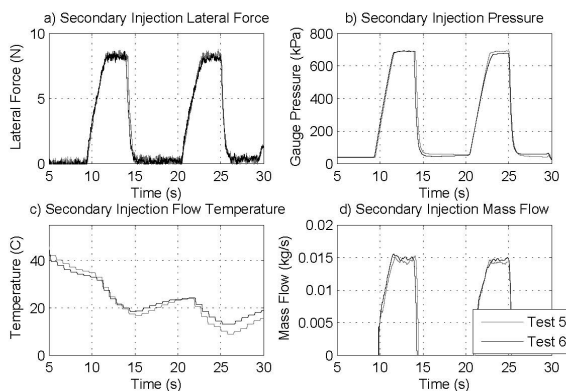


Figure 5. Side-force and secondary injection pressure for 90% injection point.

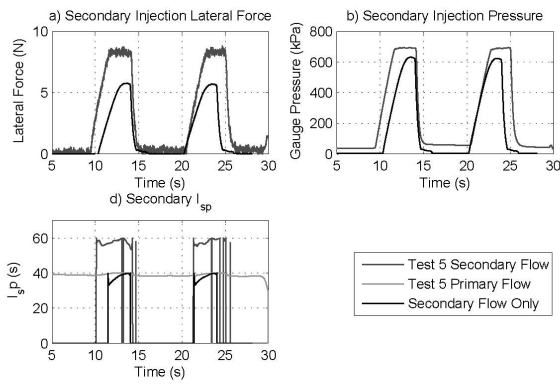


Figure 6. Side-force and secondary injection for 90% injection point for both primary flow on and secondary flow only configurations

The secondary injection orifices were machined such that they injected fluid normal to the aerospike's longitudinal axis. Lateral force, secondary injection pressure, mass flow rate, and temperatures for two typical tests are shown in Figure 5. The response fidelity between the electronic regulator control and the output size force is clearly shown. This was typical of the entire test series.

The side-force, specific impulse, and secondary flow pressure for 90% secondary injection location for both main flow on and main flow off is shown in Figure 6. The resulting side-force amplification factor and specific impulse for each configuration is shown in Table 2 and Figure 7.

Table 2. Cold flow test specific impulse results			
Test Series	Isp (s)	Isp Uncertainty (s, 95 %)	Amplification Factor
Injection location at 90%	54.8	±1.9	1.39
Injection location at 80%	47.0	±1.9	1.19
Injection location at 20%	21.2	±1.7	0.54
Secondary flow only	39.5	±1.8	

Here, side-force amplification factor was defined as the ratio of side force with a main axial flow to the side force generated by the secondary injection without the primary flow. This is similar to the definition used by Walker et al.⁵⁹ in thrust vectoring research at Johns Hopkins University. Using this definition instead of the other traditional definition of the ratio of side force to axial force specific impulse avoids the dependence on the arbitrary efficiency of the primary thruster. An additional configuration with a larger diameter injection orifice and at approximately 90% the length of the truncated spike was also tested to examine side-force scaling (Figure 8).

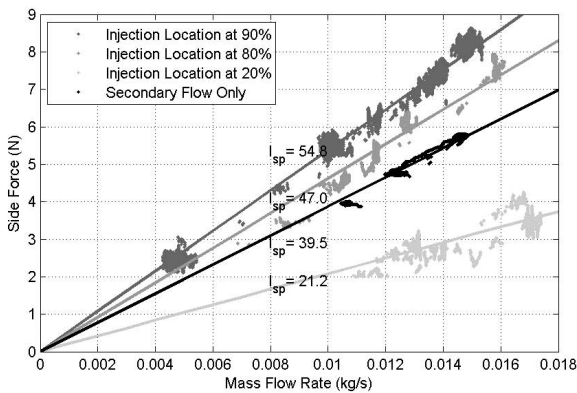


Figure 7. Cold flow secondary injection results and regressed specific impulses for various hole locations.

The use of carbon dioxide as a operating fluid leads to a useful fluid visualization technique, as carbon dioxide starts to crystallize near the end of the aerospike contour. This creates a semi-opaque white cloud that is readily visible. The temperature increase caused by shock waves resulting from secondary fluid injection create clear areas in the flow field and are thereby clearly distinguishable from the rest of the flow field. The leading edge bow shock caused by fluid injection for a high flow rate test can clearly be seen in Figure 9.

Limiting the effect of the low-pressure zone behind the injection port simply requires that the injection point be located near the end of the nozzle. This result is clearly supported by the cold flow data. For hole locations far away from the truncation length of the spike, such as the 20% injection site, the low-pressure region has more effect than the

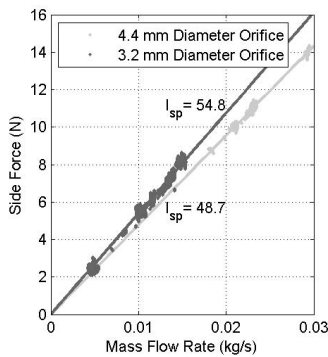


Figure 8. Cold flow secondary injection increased secondary orifice diameter results.

bow shock caused by the injection. Thus, the interaction of the main flow and the secondary fluid injection causes a side-force amplification factor less than unity. When the injection location is near the end of the aerospike, the effect of the low-pressure region is diminished, which results in large efficiency gains.

During the cold flow test series, the nozzle pressure ratio was varied from approximately 5.0 to 8.0. No meaningful correlation between

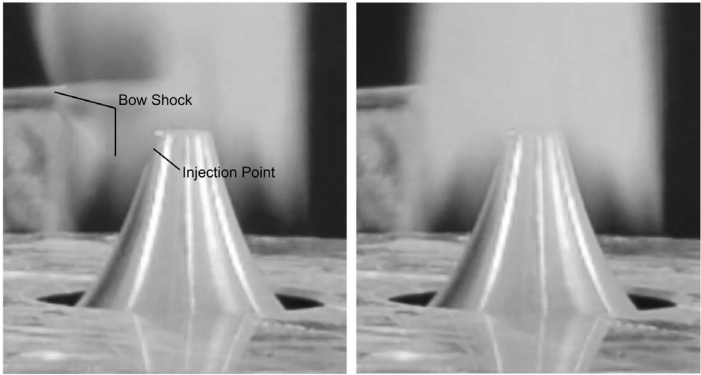


Figure 9. Aerospike cold flow test with 4.4-mm diameter orifice located at 90% of the length of the truncated aerospike. a) Thrust vectoring on, showing clear bow shock. b) Thrust vectoring off.

side force specific impulse and chamber pressure was observed over this range. The side force I_{sp} for this range with the 90% injection location is shown in Figure 10. Near the upper part of this range, the nozzle surface pressure is effectively independent of ambient pressure. At lower pressure ratios, aerospike altitude compensation will affect the local ambient Mach number and density around the secondary flow orifice. The variation of these two parameters appears to have counterbalancing influences on the side-force specific impulse over the range of pressure ratios examined during cold flow testing.

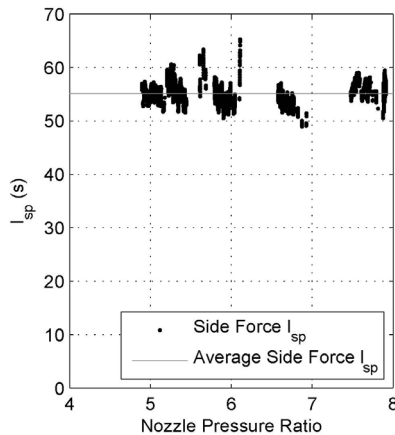


Figure 10: Cold flow specific impulse vs. nozzle pressure ratio.

It is notable that the secondary injectant does not reach sonic velocity at the immediate exit of the injection orifice. The bow shock caused by primary flow results in an effectively reduced area for the injectant immediately downstream of the orifice. This resulted in a typical drop in discharge coefficient of about 5% between tests with secondary injection only and secondary injection with active primary flow.

For aerospike configurations with the secondary injection point near the end of the aerospike, the effect of fluid injection on axial thrust was small enough such that it was not detectable by the current testing apparatus.

*II.D. Dependence on longitudinal injection site**

The high dependence of thrust vectoring efficiency with longitudinal hole location seen during this test series is in direct contradiction to side-force relations obtained on conical nozzles. For lab scale tests on conical nozzles, the optimum injection point for gas injection has been found to be nearest the throat, where the resulting bow shock does not impinge on the opposite nozzle wall.⁶⁰ For the series of cold flow tests discussed above, the optimal injection location was at the aft edge of the truncated aerospike length. Two possible explanations for this effect are the effect of local free stream Mach number at the injection location and diminishing the effect of the low-pressure, overexpanded region directly downstream of the injection location.

Several of the prominent theoretical models for thrust vectoring due to fluid injection both predict a strong positive dependence on free stream Mach number. One of these models is the blast wave theory analogy promoted by Broadwell.⁶¹ Broadwell noted the similarity between the shock wave shape exhibited by secondary injection into supersonic flow and the shock waves generated by linear charges. This

*

Nomenclature

ΔA = area of intersection for high pressure region behind shock wave

\dot{m}_j = secondary injection mass flow rate

A_j = secondary injection area

C = adjustable flow constant

F_i = amplified side force

F_v = side force without amplification by primary flow

h = accommodation height

M_∞ = free stream Mach number

$M_{j\infty}$ = Mach number of injectant downstream of injection site

P_2 = static pressure downstream of conical shock

P_∞ = static pressure of primary flow

P_s = static pressure in separated flow region

V = velocity of secondary injectant

V_j = injectant velocity

X = distance between start of separation and injection point

Symbols

α = conical nozzle half angle

δ = angle between outer boundary of separated region and nozzle wall

γ = specific heat ratio

σ = nondimensional pressure in cylindrical explosion

θ = angle between conical shock wave and nozzle wall

ϵ = angle of injectant to nozzle radial vector

neglects any dependence on boundary layer effects at the injection site, which are not insignificant. Hence, this, like the other theoretical models proposed in the 1960s should be used to predict trends only and do not make good predictors of actual side-force amplification. The defects of theoretical models for side-force injection was examined in detail by Guhse.⁶² With the aforementioned note, the blast wave theory model predicts that side-force amplification should generally obey the relation

$$\frac{F_i}{F_v} = \frac{C \sigma(\gamma) M_\infty V_\infty}{(V_j)_v} \left[1 + \frac{2 + (\gamma - 1) M_\infty^2}{2(\gamma - 1) M_\infty} \right] \quad (1)$$

which has a strong positive dependence on Mach number. Walker et al.⁵⁹ modified a linear flow relation by Vinson et al.⁶³ that assumed that the side force could be approximated by linear supersonic flow analysis over the area displaced by a secondary flow injectant expanded to the pressure of the primary flow. Their relation, once solved for the amplification factor defined above, yields

$$\frac{F_i}{F_v} = \frac{1}{(1 + \gamma)} \frac{\gamma M_\infty^2}{(M_\infty^2 - 1)^{\frac{1}{2}}} \left[\frac{1 + \gamma}{M_{j\infty}^2 [2 + (\gamma - 1) M_{j\infty}^2]} \right] \quad (2)$$

which is also strongly dependent upon the free stream Mach number. Another prominent model which seems to reproduce trends for secondary injection for conical nozzles is the relation proposed by Wu et al.⁶⁴ Their method uses conical shock analysis to approximate the pressure distribution behind the primary bow shock with a parabolic fit. They adjust for geometry but presume that the effects of the bow shock and the overexpansion region behind the injection point cancel downstream of the injection orifice. In their model, side force can be determined from the relation

$$F_i = \left[\left(\frac{P_2}{P_\infty} - 1 \right) (\Delta A - Xh) + \left(\frac{P_1}{P_\infty} - 1 \right) \left(Xh - \frac{A_j}{2} \right) \right] P_\infty \cos(\alpha) + P_\infty A_j \cos(\varepsilon) \left(\frac{P_1}{P_\infty} - 1 \right) + \dot{m}_j V_j \quad (3)$$

where

$$h = \left[\frac{2A_\infty}{\pi} \right]^{\frac{1}{2}} \left[\frac{2\gamma^2 M_\infty^2 (1 + \frac{\gamma-1}{2} M_\infty^2)}{\left(\frac{A_\infty}{A_j} \right)^2 \left(\frac{P_\infty}{P_j} \right)^2 \left(\frac{P_\infty}{P_\infty} - 1 \right) [\gamma + 1 + (\gamma-1) \frac{P_\infty}{P_\infty}]} \right]^{\frac{1}{2}} \quad (4)$$

$$X = h [\cot(\delta) + \tan(\alpha + \varepsilon)] \quad (5)$$

$$\Delta A = h^2 [\cot(\delta) + \tan(\alpha + \varepsilon)]^2 \tan \theta \quad (6)$$

The dependence of this relation on primary flow Mach number is not immediately clear. The primary Mach number is buried in the free stream static pressure, P_∞ . Ultimately, this relation results in a decreased predicted effectiveness as the injection location moves downstream towards the nozzle exit. They propose that the optimum injection location is the point farthest forward where the shock wave created by injection does not reflect off of the opposite side of the conical nozzle. Experiments conducted at the Massachusetts Institute of Technology agreed with this result.⁶⁰ The dependence of all three of these relations for a specific heat ratio of 1.2 and normalized to equal unity at a Mach number of 2.0 is shown in Figure 11.

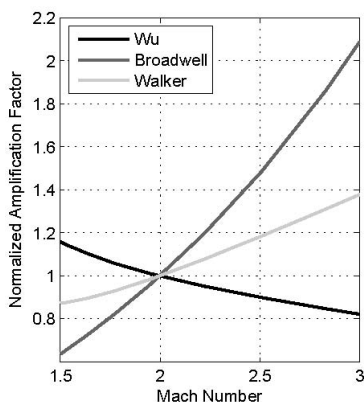


Figure 11. Normalized trends for dependence of side-force amplification on primary flow Mach number for several theoretical models and a specific heat ratio of 2.0.

The wide divergence in trends predicted by these models tends to limit confidence in their use as prediction tools; however, the model proposed by Wu et al. does seem to agree somewhat with experimental data concerning injection sites. If it is presumed that there is a positive relationship between side-force amplification and free stream Mach number, as in the models by Broadwell and Walker et al., this could explain at least some of the axial dependence measured during the aerospike cold flow test series. The free stream Mach number on the spike surface changes from approximately 1.5 at the 20% hole location to approximately 2.0 near the 80% injection location; however, because of altitude compensation, the mach number between the 80% and 90% hole locations varies only slightly. Hence, another explanation must be

sought to explain the differences in efficiency between these two locations.

An alternative insight into the cause of the high dependence on longitudinal hole location can be gleaned from secondary flow experiments performed on flat plates. For a flat plate, gaseous secondary injection creates a detached shock wave in front of the injection orifice and a low-pressure region behind caused by overexpansion of the primary injectant.^{62,65} Thus, the region directly downstream of the injection hole counteracts the force imparted by the secondary injection. For a bell nozzle, the concavity of the surface where the secondary fluid is injected tends to extend the influence of the leading shock wave. Figure 12 clearly shows the high- and low-pressure regions around the injection orifice. The effects of the leading shock wave and the low-pressure region due to overexpansion tend to cancel out in a bell nozzle.⁶⁴ Because of the convex surface of an aerospike, however, the effect of the leading bow shock tends to diminish away from the leading edge of the hole. Thus, downstream of the injection site, the net contribution to side force counteracts the secondary injection. The net side force can therefore be increased by decreasing the allowable area for the low pressure region behind the injection site. As the aerospike is sufficiently overexpanded that the base compensates to atmospheric pressure, the secondary injection entrains flow from the base region without causing an significant reduction in base pressure. Future computational results will be required to examine whether aerospike effectiveness is at all compromised in low ambient pressure conditions when the aerospike base is independent of outside pressure.

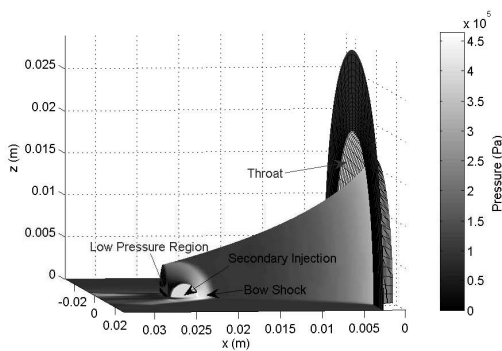


Figure 12. Bow shock and resulting static pressure from computational results on a truncated nozzle with 90% secondary injection location.

II.E. Comparison of test results with thrust vectoring effectiveness in conical nozzles

Thrust vectoring efficiency for gas injection into conical nozzles has been well established, although data for gaseous injection is not as available as that for liquid injection. Work performed by Gunter and Farenholz⁶⁰ on cold flow tests with a conical nozzle reached amplification factors of approximately 2.0. Walker et al.⁵⁹ also performed cold flow tests, including some with carbon dioxide as a working fluid, and had side-force amplification factors that ranged from approximately 1.8 to 3.0, with the highest amplification factors gained by the smallest orifices. Inouye⁶⁶ performed a series of hot gas injection tests and produced amplification factors generally between about 1.2 and 1.8 for a motor and secondary injection motor using red fuming nitric acid and unsymmetrical dimethylhydrazine.

Although the amplification factors generated for the cold flow aerospike in this test are somewhat lower than for conical nozzles, it should be noted that all of these test series involved a much higher primary flow pressure ratio than those examined in the cold flow aerospike tests for this test series. Additionally, the high-end amplification factors generated for conical nozzles generally corresponded to very small secondary orifice diameters. It is expected that variation of orifice size on an aerospike nozzle would likewise show a maximum at some orifice diameter.

In our opinion, the primary gain from thrust vectoring on an aerospike nozzle is the ability to use secondary injection jets as standalone reaction control without use of the primary engine. When the primary thruster is fired the additional benefits of flow amplification would be gained. A jet internal to a conventional nozzle would obviously not share this same operational advantage. This, coupled with the volumetric efficiency gains of aerospike nozzles makes aerospike nozzles with thrust vectoring a strong option for small satellite missions.

III. Computational Analysis

Computational analysis was performed with the Viscous Upwind Algorithm for Complex Flow Analysis⁶⁷ (VULCAN) code developed primarily by NASA Langley Research Center. VULCAN is a structured grid solver designed for supersonic flow that will intelligently neglect diffusion terms in the flow field where diffusion terms have little effect. The VULCAN solver has shown a high degree of resilience in generating second-order spatially accurate solutions to 3 or 4 orders of magnitude, a feat that prominent commercial codes have difficulty

with on flow field with supersonic separation regions such as an aerospike nozzle.

III.A. Computational Set-up

Structured grids of half an aerospike were created with a combination of NASA's Chimera Grid Tools and 3DMAGGS codes. The computational results presented here have approximately 200,000 nodes. The main throat inlet boundary condition was specified as fixed sonic flow. The secondary inlet boundary condition was specified as a subsonic inlet with a set pressure, a condition that VULCAN will solve to yield an appropriate sonic condition at the secondary flow outlet. Pressure outlet boundary conditions were used for all the grid exits.

A K-Omega turbulence model has been shown to yield reasonable agreement with data from two dimensional injection into a supersonic flow.⁶⁸ Sutherland's law was used for viscosity. The operating fluid, carbon dioxide, was modeled as a thermally perfect gas with a specific heat of 1.36. Carbon dioxide's behavior in the temperature and pressure region experienced during this test series are not necessarily well modeled as a thermally perfect gas. Modifications will be made to better model these non ideal gas effects in later computational runs.

III.B. Computational Results

Grid resolved solutions have not yet been completed; however, notional and qualitative results for computational results achieved to this point should be valid. Side-force specific impulse shown for 80% and 90% hole locations are shown in Figure 13. The results for the 90%

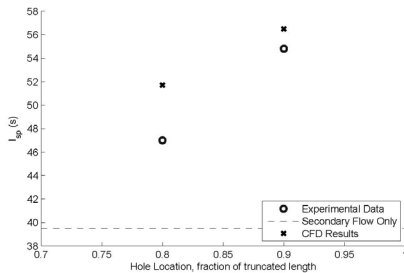


Figure 13. Side-force specific impulse vs. hole location for experimental data and computational results.

hole location show reasonable agreement with the experimental results, however, the 80% hole location, albeit following the correct trend, shows substantial divergence.

Streamlines down the plane of symmetry of the secondary flow injection port are shown in Figure 14.

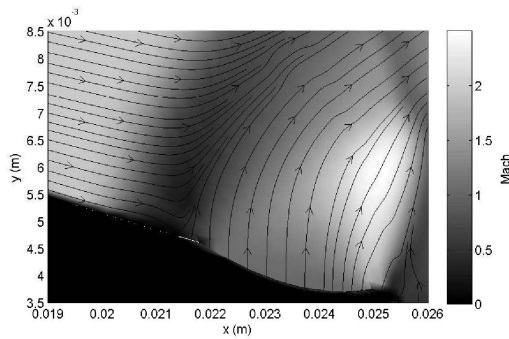


Figure 14. Stream field and Mach number near secondary injection hole for 90% injection location.

The bow shock forward of the injection hole is clearly visible, although the separated flow region does not appear to be resolved. The complex flow pattern at the base of the aerospike is shown in Figure 15.

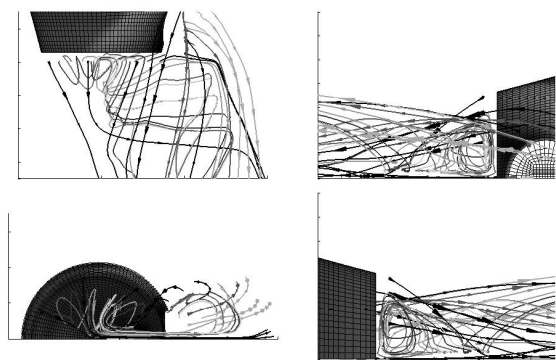


Figure 15. Three-dimensional streamlines near end of space and injection location for aerospike with 90% injection location.

IV. Conclusion

To enhance the thrust vectoring effectiveness of side-force injection on a three-dimensional aerospike nozzle, the injection site must be moved aft so that flow overexpansion does not occur on the surface of the physical spike surface. This assertion is in direct contrast to what was previously known about side injection on conventional nozzles. Data were collected for configurations with side injection port locations at 80% and 90% of the nozzle length; and significant force amplification factors were observed. The side-force specific impulse at the 90% port location is enhanced by nearly 40%. The enhanced side force Isp means that the same control impulse can be achieved for significantly less propellant.

References

- 1.. "Final Report. Studies of Improved Saturn V Vehicles and Intermediate Payload Vehicles," Tech. rep., The Boeing Company Space Division, 1996.
2. Bendersky, C., "Space Shuttle Propulsion Issue, Staged Combustion Bell Versus Tap-OFF or Gas-Generator Aerospike." Bellcomm, Inc., 1970
3. Berman, K. and Crimp Jr., F.W.C., _"Performance of Plug-Type Rocket Nozzles," ARS Journal, 1961, pp. 18-23.
4. Hendershot, K.C., Sergeant, R.J., and Wilson, H.B., "A New Approach for Evaluating the Performance and Base Environment Characteristics of Nonconventional Rocket Propulsion Systems," AIAA, Vol. AIAA-67-256, 1967, pp. all.
5. "Final Report, Advanced Aerodynamic Spike Configurations, Volume 1," Tech. rep., Rocketdyne Advanced Projects, 1967.
6. "Final Report, Advanced Aerodynamic Spike Configurations, Volume 2," Tech. rep., Rocketdyne Advanced Projects, 1967.
7. Korte, J.J., Salas, A.O., Dunn, H.J., Alexandrov, N.M., Follett, W.W., Orient, G.E., and Hadid, A.H., "Multidisciplinary Approach to Aerospike Nozzle Design," Tech. rep., National Aeronautics and Space Administration, Langley Research Center, 1997.

8. Booth, T., Vilja, J.O., Cap, D.P., and McGill, R.J., "The Design of Linear Aerospike Thrust Cells," AIAA, Vol. AIAA-1993-2562, 1993, pp. all.
9. Erickson, C., "Thrust Vector Control Selection in Aerospike Engines," AIAA Journal, Vol. 97-3307, 1997, pp. 1–6.
10. Heald, D.A. and Hart, D.A., "Advanced Reusable Engine for SSTO," 27th AIAA/SAE/ASME Joint Propulsion Conference, 1991.
11. Korte, J.J., "Parametric Model of an Aerospike Rocket Engine," Tech. rep., NASA-AIAA, 2000.
12. Higdon, K. and Landrum, D.B., "Analysis of Annular Plug Nozzle Performance and TVC," AIAA paper, Vol. AIAA-2003-4908, 2003, pp. All.
13. Ruf, J. and McDaniels, D., "Experimental Results for an Annular Aerospike with Differential Throttling," 5th International Symposium on Liquid Space Propulsion, 2003.
14. Bui, T., Murray, J., Rogers, C., Bartel, S., Cesaroni, A., and Dennett, M., "Flight Research of an Aerospike Nozzle Using High Power Solid Rockets," 41st AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, 2005.
15. Shark, S.C., Dennis, J.D., and Villarreal, J.K., "Experimental Performance Analysis of a Toroidal Aerospike Nozzle Integrated with a N2O/HTPB Hybrid Rocket Motor," 46th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, 2010.
16. Stoffel, J.R., "Experimental and Theoretical Investigation of Aerospike Nozzles in a Hybrid Rocket Propulsion System," 47th AIAA Aerospace Sciences Meeting, 2009.
17. Lemieux, P., "Development of a Reusable Aerospike Nozzle for Hybrid Rocket Motors," 39th AIAA Fluid Dynamics Conference, 2009.
18. Lemieux, P., "Nitrous Oxide Cooling in Hybrid Rocket Nozzles," Progress in Aerospace Sciences, Vol. 46, 2009, pp. 106_115.
19. Ladeinde, T. and Chen, H., "Performance Comparison of a Full-

Length and a Truncated Aerospike Nozzle,” 46th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, 2010.

20. Besnard, E., Chen, H.H., and Mueller, T., “Design, Manufacturing and Test of a Plug Nozzle Rocket Engine,” AIAA paper, Vol. AIAA-2002-4038, 2002, pp. all.

21. Besnard, E. and Garvey, J., “Aerospike Engines for Nanosat and Small Launch Vehicles (NLV/SLV),” Space 2004 Conference & Exhibit, 2004.

22. Besnard, E. and Garvey, J., “Development and Flight-Testing of Liquid Propellant Aerospike Engines,” 40th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, 2004.

23. Wilson, A., Clark, J., Besnard, E., and Baker, M., “CFD Analysis of a Multi-Chamber Aerospike Engine in Over-Expanded, Slipstream Conditions,” 45th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, 2009.

24. Schoyer, H.F.R., “Thrust Vector Control for (Clustered Modules) Plug Nozzles: Some Considerations,” Journal of Propulsion and Power, Vol. 16, 2000, pp. 196–201.

25. Hallard, R. and Merienne, M.C., “Aerospike Nozzle Tests,” Proceedings of the Third European Symposium on Aerodynamics for Space Vehicles, Vol. 426, 1998, pp. 387_394.

26. Fick, M., “Performance Modeling and Systems Aspects of Plug Cluster Nozzles,” 34th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, 1998.

27. Fick, M. and Schmucker, R.H., “Remarks on Plug Cluster Nozzles,” 31st AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, Vol. AIAA 95-2694, 1995.

28. Hagemann, G., Schley, C.-A., Odintsov, E., and Sobatchkine, A., “Nozzle Flowfield Analyses With Particular Regard to 3D-Plug Cluster Configurations,” AIAA, Vol. AIAA-1996-2954, 1996, pp. all.

29. Hagemann, G., Immich, H., Nguyen, V., and Dumnov, G., “Advanced Rocket Nozzles,” Journal of Propulsion and Power, Vol.

14, No. 5, September–October 1998, pp. 620–634.

30. Rommel, T., Hagemann, G., Schley, C., Manski, D., and Krulle, G., “Plug Nozzle Flowfield Calculations for SSTO Applications,” 31st Joint Propulsion Conference and Exhibit, AIAA 95-2784, 1995.

31. Sorge, R., Carmicino, C., and Nocito, A., “Design of a Lab-Scale Cooled Two-Dimensional Plug Nozzle for Experimental Tests,” 38th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, 2002.

32. Nasuti, F., Geron, M., and Paciorri, R., “Three Dimensional Features of Clustered Plug Nozzle Flows,” 39th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, 2003.

33. Nasuti, F. and Onofri, M., “Analysis of In-Flight Behavior of Truncated Plug Nozzles,” 36th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, 2000.

34. Onofri, M. and Nasuti, F., “Prediction of Open and Closed Wake in Plug Nozzles,” Proc. 4th Europ Symp. Aeroghermodynamics for Space Applications, 2002.

35. Onofri, M., “Plug Nozzles: Summary of Flow Features and Engine Performance,” AIAA, Vol. 1-007, 2002, pp. 1–26.

36. Fujii, K., Imai, K., and Sato, T., “Computational Analysis of the Flow Field Near the Boat-Tail Region of Annular Plug Nozzles,” JSME International Journal, Vol. 45, 2002, pp. 745–751.

37. Ito, T. and Fujii, K., “Flow Field and Performance Analysis of an Annular-Type Aerospike Nozzle with Base Bleeding,” Transactions of the Japan Society for Aeronautical and Space Sciences, Vol. 46-151, 2003, pp. 17–23.

38. Ito, T. and Fujii, K., “Numerical Analysis of the Base Bleed Effect on the Aerospike Nozzles,” Transactions of the Japan Society for Aeronautical and Space Sciences, Vol. 46-151, 2003, pp. 17–23.

39. Ito, T. and Fujii, K., “Flow Field Analysis of the Base Region of Axisymmetric Aerospike Nozzles,” 39th AIAA Aerospace Sciences Meeting & Exhibit, No. AIAA 2001-1051, 2001.

40. Ito, T., Fujii, K., and Hagemann, G., "Numerical Investigation of the Side-Fence Effect on Linear Plug Nozzle Performance," AIAA, Vol. AIAA 2004-4018, 2004, pp. all.
41. Ito, T., Fujii, K., and Hayashi, A., "Computations of the Axisymmetric Plug Nozzle Flow Fields: Flow Structures and Thrust Performance," 17th AIAA Applied Aerodynamics Conference, No. AIAA 1999-3211, 1999.
42. Miyamoto, H., Matsuo, A., and Kojima, T., "Effects of Sidewall Configurations on Rectangular Plug Nozzle Performance," AIAA, Vol. AIAA-2006-4373, 2006, pp. all.
43. Negishi, H. and Fujii, K., "Computational Analysis of the Effective Secondary-Flow Injections for the Plug-Nozzle Drag Reduction," 33rd AIAA Fluid Dynamics Conference and Exhibit, 2003.
44. Tsutsumi, S., Yamaguchi, K., Teramoto, S., and Nagashima, T., "Clustering Effects on Performance and Heating of Linear Aerospoke Nozzle," 45th AIAA Aerospace Sciences Meeting & Exhibit ; 8-11 January 2007, Reno, Nevada, 2007.
45. Taniguchi, M., Mori, H., Nishihira, R., and Niimi, T., "Experimental Analyses of flow Field Structures around Clustered Linear Aerospoke Nozzles," American Institute of Physics Conference Proceedings, Vol. 762, 2005, pp. 349–354.
46. Tomita, T., Takahashi, M., Onodera, T., and Tamura, H., "A Simple Performance Prediction Model of Clustered Linear Aerospoke Nozzles," 37th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, 2001.
47. Tomita, T., Kumada, N., and Ogiwara, A., "A Conceptual System Design for a Linear Aerospoke Engine Applied to a Future SSTO Vehicle," 46th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, 2010.
48. Tomita, T., Takahashi, M., Onodera, T., and Tamura, H., "Effects of Base Bleed on Thrust Performance of a Linear Aerospoke Nozzle," 35th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, No. AIAA 1999-2586, 1999.
49. Chang-Hui, W., Liu, Y., and Qin, L.-Z., "Aerospoke nozzle contour

design and its performance validation,” *Acta Astronautica*, Vol. 64, 2009, pp. 1264–1275.

50. Chun-Guang, J., Yu, L., Chang-Hui, W., Wen-Bo, X., and Zhen, L., “A Study for Thrust Vector Control of Aerospike Nozzle Based on Second Injection,” *Journal of Propulsion Technology*, Vol. 30-1, 2009, pp. 66–71.

51. Spring, J., Xiao, P., Yu, L., and Yunfei, L., “Plug Nozzle Thrust Vector Control Study,” *Technological Sciences*, Vol. 3, 2009, pp. 505–510.

52. Li, J., Liu, Y., Liao, Y., Wang, C., Wang, Y., and Wang, N., “Experimental and Numerical Study on Two Dimensional Plug Nozzle,” 46th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, 2010.

53. Yu, L., Wuye, D., Zhengke, Z., Lizi, Q., and Yibai, W., “Numerical Investigation on Linear Aerospike Nozzles,” 37th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, 2001.

54. Karthikeyan, N., Verma, S. B., and Venkatakrishnan, L. V., “Experimental Investigation of the Acoustics of an Annular Aerospike Nozzle Flow,” 15th AIAA/CEAS Aeroacoustics Conference (30th AIAA Aeroacoustics Conference), 2009.

55. Verma, S. B., “Performance Characteristics of an Annular Conical Aerospike Nozzle with Freestream Effect,” 44th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, 2008.

56. Kraiko, A. N. and Tillyayeva, N. I., “Contouring Spike Nozzles and Determining the Optimal Direction of their Primary Flows,” *Fluid Dynamics*, Vol. 42-2, 2007, pp. 321–329.

57. Naghib-Lahouti, A. and Tolouei, E., “Investigation of the Effect of Base Bleed on Thrust Performance of a Truncated Aerospike Nozzle In OFF-Design Conditions,” *European Conference on Computational Fluid Dynamics*, 2006.

58. Ruf, J. H. and McConnaughey, P. K., “The Plume Physics Behind Aerospike Nozzle Altitude Compensation and Slipstream Effect,” 33rd

AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, 1997.

59. Walker, P. E., Stone, A. R., and Shander, M., "Secondary Gas Injection in a Conical Rocket Nozzle: Effect of Orifice Diameter and Molecular Weight of Injectant," Originally created at Johns Hopkins University; Reproduced by the Armed Services Technical Information Agency (now called Defense Technical Information Center).

60. Gunter, F. L. and Farenholz, F. E., "Final Report on a Study of Rocket Thrust Control by Gas Injection," Tech. rep., Massachusetts Institute of Technology Naval Supersonic Laboratory, 1961.

61. Broadwell, J. E., "Analysis of the Fluid Mechanics of Secondary Injection for Thrust Vector Control," AIAA Journal, Vol. 1, 1963, pp. 1067–1075.

62. Guhse, R. D., "On Secondary Gas Injection in Supersonic Nozzles," Journal of Spacecraft and Rockets, Vol. 3, 1966, pp. 143–149.

63. Vinson, P. W., Amick, J. L., and Liepman, H. P., "Interaction Effects Produced by Jet Exhaustion Laterally Near Base of Ogive-Cylinder Model in Supersonic Main Stream," Tech. rep., NASA and University of Michigan, 1959.

64. Wu, J.-M., Chapkins, R. L., and Mager, A., "Approximate Analyses of Thrust Vector Control by Fluid Injection," ARS Journal, 1961, pp. 1677–1684.

65. Zukoski, E. E. and Spaid, F. W., "Secondary Injection of Gases into a Supersonic Flow," AIAA Journal, Vol. 2, 1964, pp. 1689–1697.

66. Inoyue, T., "Experiments on Rocket Thrust Vector Control by Hotgas Injection," Journal of Space, Vol. 3-4, 1966, pp. 737–739.

67. VULCAN Home Page, <http://vulcan-cfd.larc.nasa.gov/> [retrieved 30 November 2010].

68. Zukoski, E., Sapid, F., Clarence, C., Philips, B., Sriram, A., and Matthew, J., Injection into a Supersonic Stream, Fluent, Inc, 2005.

Closed-Loop Attitude Control Using Fluid Dynamic Vectoring on an Aerospike Nozzle

Nathan Erni and Crystal Frazier
Utah State University

Abstract

A control of a prototype satellite bus using fluid mechanical vectoring on an aerospike nozzle is demonstrated. The design achieves thrust vectoring by injecting propellant asymmetrically into the unconstrained aerospike exhaust plume near the nozzle base. The prototype system uses cold-gas thrusters to both spin-up and de-tumble the test article. The system is configured with axially directed annular flows that produce large longitudinal thrusts and smaller secondary lateral injection flows for side thrusts. Both open- and closed-loop attitude control, with and without main aerospike annular flow active, are demonstrated. Proportional, integral, derivative (PID) regulation is used for closed-loop attitude control. When the main annular flow is active, the achieved side forces with secondary injection are significantly greater than can be achieved by direct reaction control alone. When the vectoring ports are operated with no primary plenum flow, very small 'impulse bits' are generated. Based on the results presented in this paper, a significant potential exists for three-degree-of-freedom (3-DOF) atti-

tude control without mechanical nozzle gimbals. When extended to 3-DOF, the closed-loop control law will allow the primary satellite propulsion system to be used for both larger-scale orbit change maneuvers and smaller-scale proximity operation maneuvers with the same system.

I. Introduction

During the past decade, advances in miniature electronics and exponential growth in computational power have allowed the development of powerful satellites with a very small-scale form factor. Within the next decade, these small spacecraft, referred to as ‘CubeSats,’ will have the capability to perform a variety of in-space missions that previously could only be performed by very expensive, large-scale satellites. A wide variety of missions from science and exploration and telecommunications missions to military reconnaissance and surveillance (R&S) missions would be enabled by a constellation of CubeSats precisely positioned to achieve a strategic objective. Launched as a constellation, this distributed ‘swarm’ offers distinct advantages not achievable by single, larger-scale spacecraft.

For military R&S applications, a constellation allows several member elements to fail and still achieve mission objectives. The multiple spacecraft allow for simultaneous spatial and temporal measurements once the target destination is achieved. A large constellation is multiply redundant and would be nearly impossible to disable with airborne or ground-based weapon systems. Disabling a single member of the constellation still leaves a system with 95–99% functional capability. Most importantly, a highly dispersed orbiting constellation would enable the U.S. Air Force and the National Reconnaissance Office to counter many enemy denial and deception practices. Such a constellation of spacecraft would provide virtually undeniable critical intelligence for multiple purposes, including detecting, locating, characterizing, and tracking 1) weapons of mass destruction and their support and production infrastructure; 2) narcotics and terrorism activities, organizations and leadership; 3) potential adversary’s air, land, and sea military activities; 4) air and ground moving targets and single entities; and 5) advanced weapons systems.

For civilian space missions, providing a capability of approximately 800 m/sec allows the proposed constellation to be launched as a ‘ride-along’ payload and deployed onto interplanetary trajectories from a standard geostationary transfer orbit. Accomplishing interplanetary missions as a secondary payload for commercial geosynchronous trans-

fer orbit launches offers a potentially game-changing technology for National Aeronautics and Space Administration (NASA) science. Once the target destination is achieved, the distributed nature of this ‘interplanetary swarm’ will allow simultaneous spatial and temporal measurements. A single large spacecraft cannot achieve such measurements.

Because of their small sizes, CubeSats must be constructed using the most efficient packaging possible. Thus, the design challenges associated with creating CubeSat-scale propulsion systems are greater than those associated with designing thrusters for conventional spacecraft. Deploying conventional propulsion systems with gimballed bell-nozzles for attitude control is infeasible in such small form factors. This project researches nonstandard propulsion systems designs that allow efficient packaging and provide high impulse levels for orbital maneuvers and still allows small impulse bits for orbital rendezvous and precision formation flying.

The aerospike nozzle is a key enabling technology to be investigated here. An aerospike nozzle is an altitude-compensating nozzle that maintains aerodynamic efficiency across various altitudes. Because of the unconstrained external plume, it is possible to achieve thrust vectoring fluid-dynamically by injecting propellant asymmetrically at points along the nozzle contour. If this vectoring potential can be harnessed and incorporated into a closed-loop thrust-vectoring scheme, there exists a significant potential for three-degrees-of-freedom (3-DoF) attitude control without mechanical nozzle gimbals. The result is a significant reduction of system complexity and a significant reduction of the overall propulsion system weight. Finally, when the vectoring ports are operated with no primary plenum flow and in a pulse-width mode, very small ‘impulse bits’ can be generated, which offers the potential for the system to also be used for precise proximity operations. This small impulse option is not possible with a conventional nozzle.

Despite many potential advantages over conventional nozzle designs, the aerospike nozzle has never been deployed on an operational vehicle. This ‘deployment gap’ is due to (justifiably or not) the perceived low overall technology readiness level (TRL) of the aerospike configuration. This low TRL is especially relevant with regard to very small thrusters on a scale useful for CubeSat applications. A single-axis prototype has been developed to demonstrate the proposed technology. Results of this demonstration project will be presented herein. A primary goal of the demonstration project is the rapid upgrade of the technology readiness level of the system from the current level of approximately 2–3 to a level of approximately 4. Further upgrading the system TRL to 6 by follow-on development efforts would allow an aerospike-based propulsion system with closed loop vectoring control

to be seriously considered as a viable alternative for future space deployments.

This thesis proposes the demonstration of a novel, compact propulsion system, scaled for CubeSat-sized spacecraft. Once developed, the system will provide the capability to precisely position CubeSats to form a large constellation whose members work collectively to accomplish a meaningful tactical objective. The distributed nature of this ‘swarm’ offers distinct advantages not achievable by a single, large-scale spacecraft. However, design challenges are associated with creating small CubeSat-scaled propulsion systems compared to conventional propulsion systems with gimbaled nozzles. The design, based upon the aerospike nozzle concept, ameliorates this difficulty. These advantages will be explored in this article.

II. Objectives

The objectives, verification methods, and achievements are listed below. This research is performed in conjunction with phase one of the CubeSat-Scale Propulsion System (CCSPS), whose purpose is to minimize NASA’s deployment gap by rapidly upgrading the TRL level from the current level 3 to approximately level 4. Tests were performed by a hardware in-the-loop simulation to demonstrate, measure, and assert the control characteristics of an aerospike nozzle. The approach was to use a single aerospike nozzle with two secondary injection ports for angular control.

1. Open-Loop Thrust Vector Control (TVC) Secondary Flow Only: The satellite starts at a resting state while a single secondary injection port is turned on, causing the satellite to spin up to a desired angular rate of 2 rad/s (114.6 deg/s). The test is conducted without the presence of annular flow. This test set will show the ability of rotational ramping using an aerospike nozzle showing the ability to de-tumble a satellite without annular flow.
2. Open-Loop TVC Annular Flow and Secondary: The satellite starts at a resting state while a single secondary injection port is turned on, causing the satellite to spin up to a desired angular rate of 2 rad/s (114.6 deg/s). The test is conducted with the presence of annular flow. This test set will show the ability of rotational ramping using an aerospike nozzle showing the ability to de-tumble a satellite with annular flow.
3. Closed-Loop TVC Secondary Flow Only: The satellite starts at a resting state while the secondary injection

ports are controlled by a proportional, integral, derivative (PID) controller using pulse-width modulation. The satellite will rotate to a desired angular position. The test is conducted without the presence of annular flow. This test set will show the ability of closed-loop control using an aerospike nozzle for positioning or maneuvering a satellite without annular flow.

4. Closed-Loop TVC Annular Flow and Secondary Flow: The satellite starts at a resting state while the secondary injection ports are controlled by a PID controller using pulse-width modulation. The satellite will rotate to a desired angular position. The test is conducted with the presence of annular flow. This test set will show the ability of closed-loop control using an aerospike nozzle for positioning or maneuvering a satellite with annular flow.

III. Background

While the aerospike nozzle has long been known for its altitude compensation capability during endo-atmospheric flight, the aerospike also presents significant advantages for purely in-space applications. Because of its shape, the aerospike nozzle can be constructed with a higher area expansion ratio and more compact form factor than a conventional bell nozzle of the same mass. The higher expansion ratio provides better performance in a space environment; the compact form factor offers a very significant advantage for CubeSat-scale spacecraft where volume efficiency is a key consideration.

Figures 1 and 2 [1] compare the aerospike nozzle theory of operation to the conventional bell nozzle typically used for space applications. Here the walls of the conventional bell nozzle constrain the flow and result in an overexpanded plume for launch conditions and an underexpanded plume for in-space operating conditions. In both cases, the performance is significantly lower than optimal. The aerospike nozzle, however, does not constrain the outer boundaries of the flow slipstream. The plume is free to expand or contract depending on the external pressure at the operating conditions. Compared with the bell nozzle, the achieved performance is significantly higher for both launch and in-space operating

Compared with the 100:1 expansion ratios typically available to bell nozzles, annular aerospike nozzles (of equivalent mass) with expansion ratios exceeding 250:1 can be easily fabricated. This higher expansion ratio increases vacuum specific impulse (Isp) by more than 10%. The increase in Isp results in a 10–18% reduction in the propel-

lant mass, and both factors produce an 8–12% reduction of the total system weight.

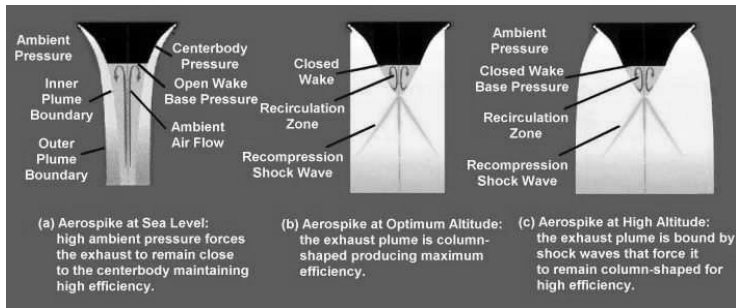


Figure 1. Aerospike operational comparison for (a) overexpanded, (b) optimally expanded, and (c) underexpanded conditions.

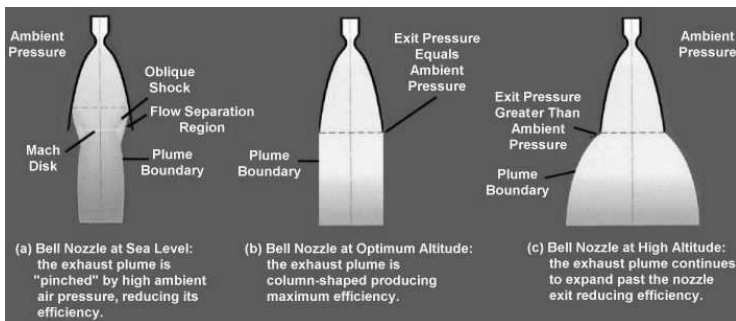


Figure 2. Conventional bell nozzle operational comparison for (a) overexpanded, (b) optimally expanded, and (c) underexpanded conditions.

IV. Literature Review

A. Prior Art

Testing on aerospike nozzles began in the 1950s in preparation for the Saturn V rocket [2]. After extensive research and test series, Rocketdyne concluded that there was no apparent benefit for the use of an aerospike over a conventional nozzle. The series of tests performed were based on liquid injection and indicated that an annular throat aerospike nozzle is at least comparable to and in many instances better than the 80% length bell nozzle at design pressure ratio [1, 3], but had less or equal thrust vectoring capability. Research continued into the

1970s as an aerospike nozzle was under consideration for the Space Shuttle's main engine [4, 5] but again was not selected. Efforts on aerospike research declined until the 1990s when NASA proposed a linear aerospike for propulsion system for the X-33 and the VentureStar [6]. These programs were later canceled.

More recently, analytical research has been conducted worldwide by several institutions for the development and performance analysis of TVC [7, 8], differential throttling [9], clustering performance [10–13], slipstream effects [14–16], base bleed injection [17–19], optimal contours [20], and acoustics [21]. Hardware experiments have been conducted by Arizona State University [22], University of Washington [23], and California Polytechnic [24]. California State University in conjunction with Garvey Spacecraft corporation has conducted several experiments along with launching several sounding rockets [25, 26].

B. Cold Flow Testing

Utah State University recently performed analytical and experimental evaluations on aerodynamic thrust vectoring on aerospike nozzles using secondary injection [27]. These experiments included sizing an aerospike nozzle for slightly above optimal expansion for the testing altitude. This design allowed compression waves to impinge past the end of the nozzle to create a pressure distribution along the spike similar to vacuum conditions. For these tests carbon dioxide (CO_2) was used as the working fluid. Different spike configurations were designed and tested for an annular mass flow rate of approximately 1 kg/s with a secondary injection flow rate between 2 and 3% of the annular mass flow. Figure 3 shows the designed spikes truncated to 57% of the theoretical spike length. This level of truncation has minimal effect on the overall nozzle performance [28].

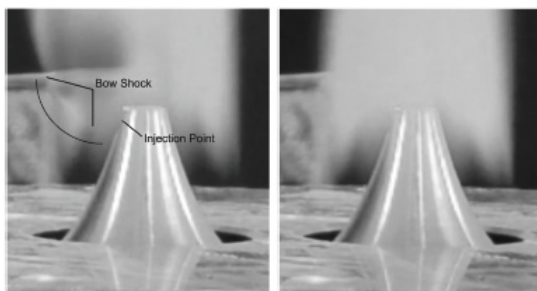


Figure 3. Aerospike cold flow test with 4.4-mm-diameter orifice located at 90% of the length of the truncated aerospike. (a) Thrust vectoring active, showing bow shock. (b) Thrust vectoring inactive.

The spike nozzle housed the secondary injection ports. Tests were performed with secondary injection ports located at 20, 80, and 90% of the truncated spike. Figure 4 shows experimental results from secondary injection tests performed with and without main plenum flow. The key feature to take away from this figure is that the generated forces with secondary flow active are greater than those generated with secondary injection only. Test results presented in Figure 5 indicate the optimal injection site is located at approximately 90% of the truncated spike. This result implies that the same control impulse can be achieved for significantly less propellant when the main plenum flow is active.

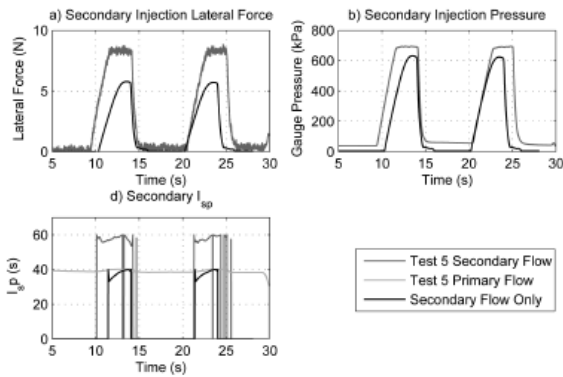


Figure 4. Side force and secondary injection for 90% injection point for both primary flow active and secondary flow only configurations.

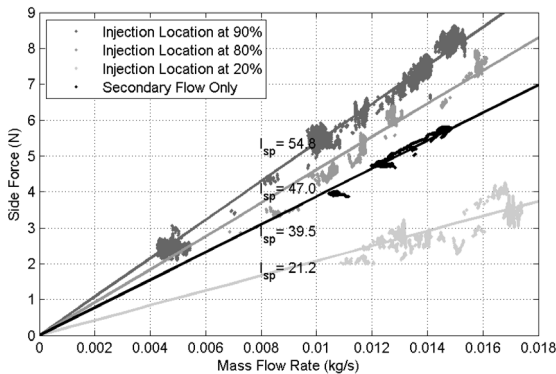


Figure 5. Cold flow secondary injection results and regressed specific impulses for various hole locations.

Table I summarizes the test results including side-force amplification factors and achieved specific impulses. Here the side force amplification is defined as the ratio of side force with a main axial flow to the side force generated by the secondary injection without the primary flow. Results show a higher amplification factor is produced as the injection location is moved closer towards the end of the truncated region.

Table I
COLD FLOW TEST SPECIFIC IMPULSE RESULTS.

Test Series	Isp(s)	Uncertainty (s, 95%)	Amplification Factor
Injection Location at 90%	54.8	1.9	1.39
Injection Location at 80%	47.0	1.9	1.19
Injection Location at 20%	21.2	1.7	0.54
Secondary Flow Only	39.5	1.8	

V. CubeSat Background

As described in the background section, the primary focus of this research is to demonstrate the feasibility for secondary injection thrust vectoring of a CubeSat using an aerospike nozzle. A CubeSat is a miniature satellite with a predefined external form factor. The size is usually described in terms of volume using "1U" as a 10-cm cube. The sizing for a typical CubeSat is shown in Figure 6.

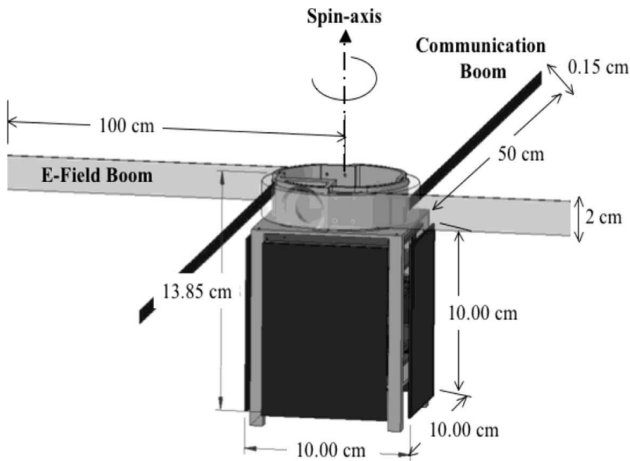


Figure 6. Typical 1U CubeSat configuration.

The current method for deploying CubeSats is the Poly- PicoSatellite Orbital Deployer (P-POD) developed by California Polytechnic State University [29]. Each P-POD is mounted as a secondary payload on the launch vehicle and carries several 1U CubeSats. Figure 7 shows a typical P-POD configuration. Once the desired orbit is reached, the P-POD releases the CubeSats. With the current state-of-the-art, size restrictions and hazards to the primary payload restrict CubeSats from having integrated propulsion systems. Thus, CubeSat orbits are determined by the launch insertion and P-POD deployment velocities. Because of its shape, the aerospike nozzle can be constructed with a higher area expansion ratio and more compact form factor than a conventional bell nozzle of the same mass. The higher expansion ratio provides better performance in a space environment; the compact form factor offers a very significant advantage for CubeSat-scale spacecraft where volume efficiency is a key consideration.

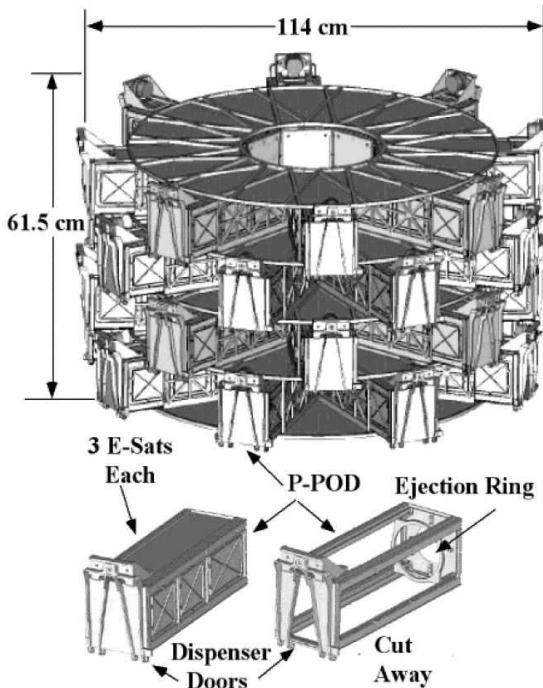


Figure 7. Proposed compact CubeSat scaled propulsion system (CCSPS).

Figure 8 shows a proposed compact CCSPS. Figure 9 shows the CCSPS integrated with the satellite bus and the P-POD dispensing system. The aerospike nozzle is housed in a fixed 10-cm diameter plenum containing hybrid propellant. The proposed design combines three emerging technologies—1) aerospike or external plume nozzles; 2) hybrid rocket systems; and 3) direct digital manufacturing—to build a unique propulsion unit that can potentially enable the constellation described in the previous paragraphs. Because the system would be flown in an inert condition and would carry no ordnance, multiple payloads can be piggy-backed together with no overall increase to mission risk. Two of these aerospike nozzles would be placed on the base of a 2U CubeSat module as shown in the aft view of Figure 9. The potential CubeSat propulsion system would be self-contained in a 2U CubeSat, housing all of the propellant, sensors, controllers, and valves. This design allows for any existing CubeSat to be attached, allowing for orbital maneuverings. Approximately 50 grams of thermo-plastic hybrid fuel and 0.7 kg of liquid nitrous oxide as the oxidizer would enable a 1-kg spacecraft to achieve a Delta V of 800 m/sec. This velocity change is sufficient to achieve escape velocity from a geosynchronous transfer orbit.

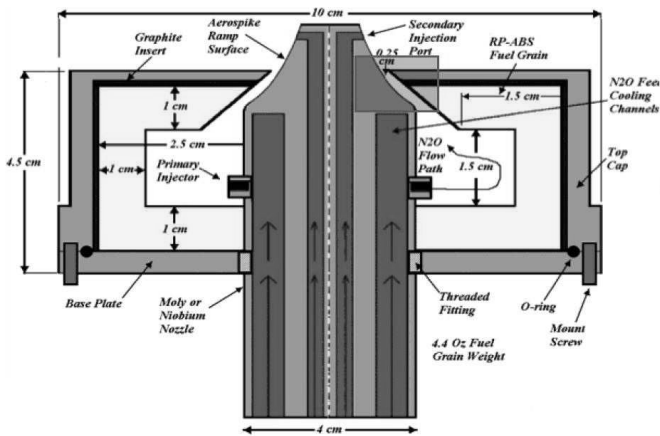


Figure 8. Regeneratively cooled, low-profile aerospike/hybrid motor.

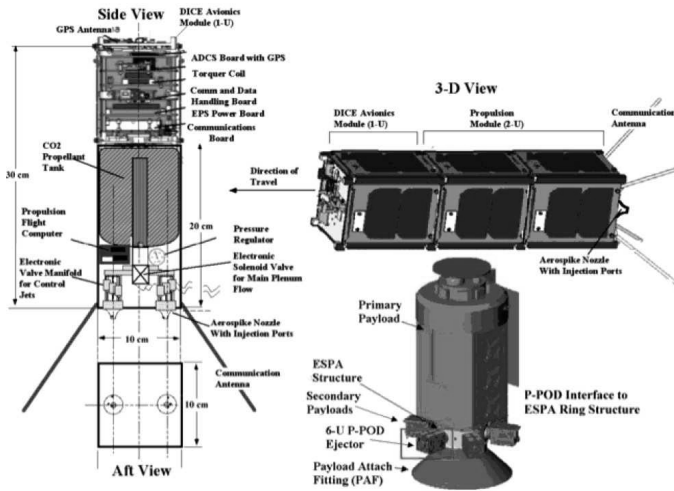


Figure 9. Aerospike configuration (forward) attached to a 1U CubeSat and (aft) inside a P-POD dispensing system.

The experimental prototype demonstrated in this research approximates the CCSPS features by using a single-axis scaled aerospike nozzle for vectoring in the yaw axis.

VI. Methodology

This chapter describes all design aspects and tests for this project, including the aerospike design, prototype satellite design, housing cage design, instrumentation, control methodology, and experimental tests.

A. Aerospike Design

The contour of the spike was calculated using the methods of Lee and Thompson [30]. Figure 10 shows the aerospike geometries. A preliminary trade study was performed to estimate the nozzle parameters and operating systems that were most advantageous to accomplishing the proposed research objectives allowing for a sufficient mass flow rate, aerospike diameter, and overall propellant. Figure 11 shows an aerospike sizing trade study including a detailed description comparing operating pressure, mass flow rate, produced thrust, plug diameter, outer throat diameter, and expansion ratio. This plot was used to calculate the aerospike's size for the given operating pressure of 1.315 kPa and

mass flow rate of around 0.1 kg/s used. Table II summarizes the resulting design features.

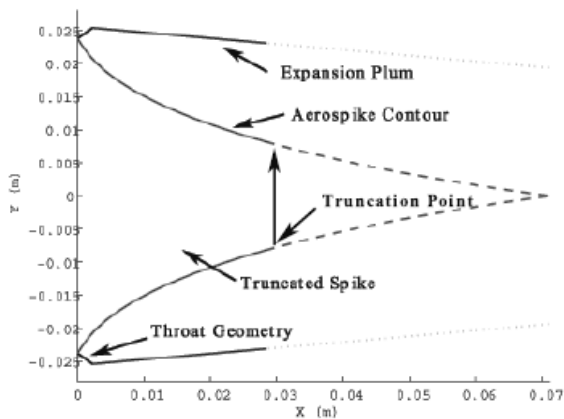


Figure 10. Aerospike nozzle contour and dimensions.

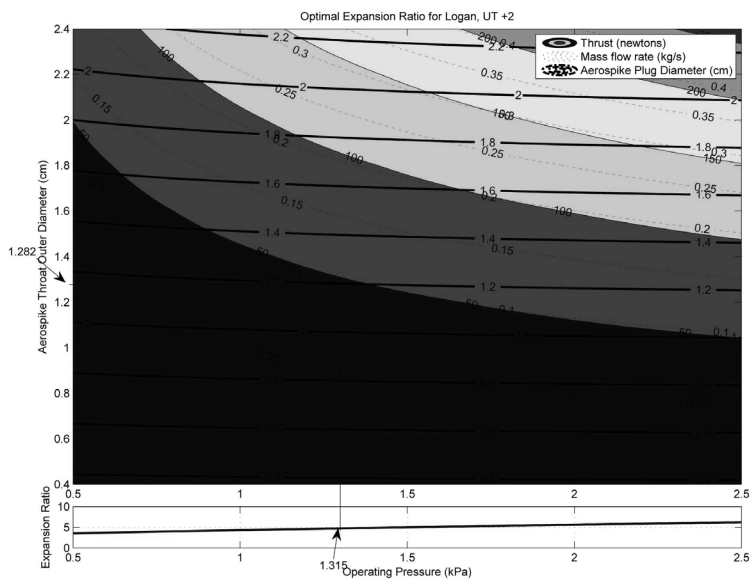


Figure 11. Aerospike sizing trade study.

Table II
AEROSPIKE PARAMETERS

Parameter	Value	Parameter	Value
Pressure Operating	1315 kPa	Gamma	1.28
Temperature	300 K	MW	44.01
Radius Cow1	0.006415	Sample Points	1500
Pressure Ambient	86.12625 kPa	Truncation	.6
Expansion Ratio	4.75	Gravity	9.8067

Figure 12 shows the design features of the prototype test article. The system was sized for an annular flow rate of 0.1 kg/s with two secondary injection ports each with an approximate flow rate of about 0.002 kg/s. The plenum volume was sized to allow CO₂ gas to equalize throughout the chamber to maintain a symmetrical flow along the aerospike.

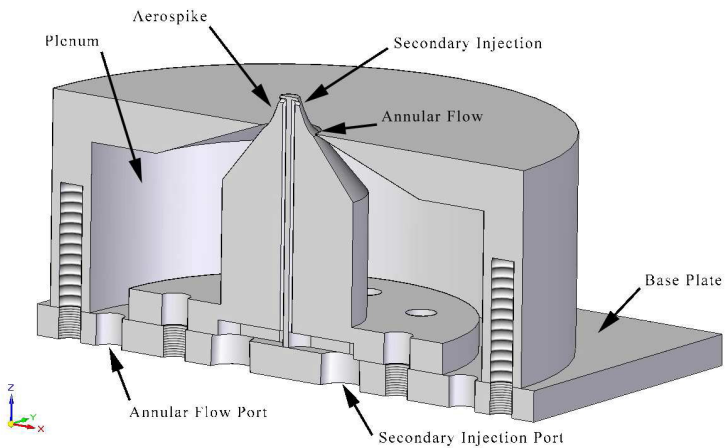


Figure 12. Computer-automated design model of the prototype aerospike configuration.

Figure 13 shows the aerospike configuration without the outer chamber attached to the base plate of the test apparatus. The annular flow port blows into a flat dispersing plate causing the flow to be uniform within the plenum.

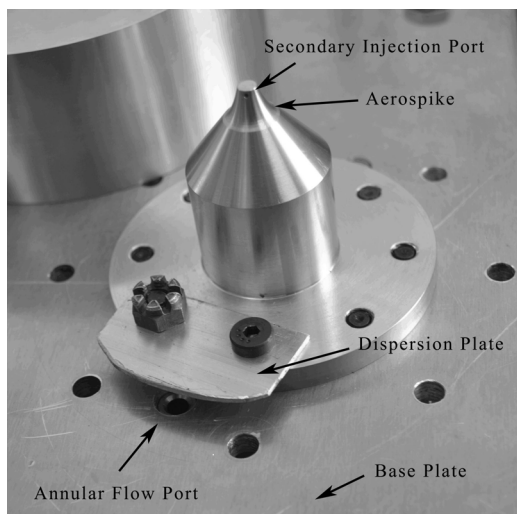


Figure 13. View of the machined aerospike with the dispersion plate for annular flow.

B. Prototype Satellite Design

Figure 14 shows the prototype apparatus designed to approximate a scaled-up CubeSat. Figure 15 shows the yaw axis of rotation. This scale was required to economically house the non-flight-weight propellant tanks and valves used to control the system. The $1' \times 1' \times 2'$ structure is supported by two aluminum base plates with aluminum rods connecting each corner. A square aluminum tubing placed in the center of the satellite supports four 24-ounce CO_2 tanks and internally houses two 14.8-VDC Li-Ion battery packs. The top of the tubing is bolted to a sheet of $1/4"$ polycarbonate, locking the CO_2 tanks in a downward position for the use of liquid propellant. All instrumentation devices are housed on the polycarbonate sheet except for the inertial measurement unit and one pressure and temperature sensor housed on the bottom of the satellite. Included in the instrumentation are DC-DC regulators, pressure and temperature sensors, pressure regulator, annular and secondary flow valves, load cell, Gumstix for control code and communication, and a data acquisition module for sensory input. Table III lists the parts with a further description of each major part. Other material not listed includes aluminum plumbing, wiring, and connectors. All communication to and from the satellite is transferred through a wireless link eliminating the need for a slip ring or excess cables.

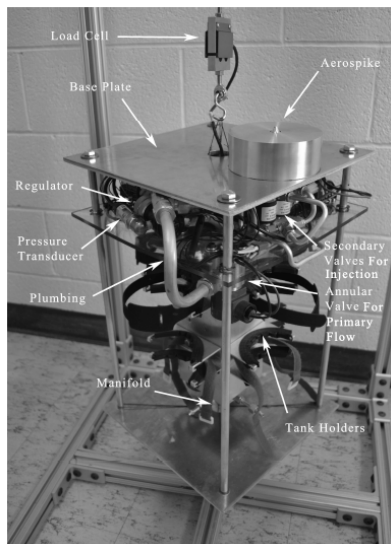


Figure 14. Prototype satellite housing design and configuration.

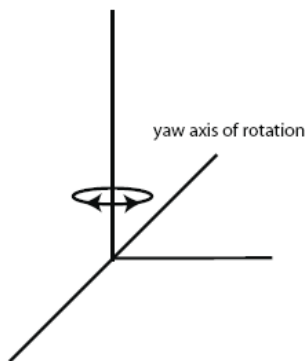


Figure 15. Prototype satellite axis of rotation.

Table III
AEROSPIKE-CONTROLLED SATELLITE SYSTEM AVIONICS PARTS LIST

Satellite Parts	Description
Inertial Measurement Unit (IMU)	Micro-Strain's 3DM-GX3-35
Customized Polymer Li-Ion Battery	14.8 VDC 6400 mAh (quantity of 2)
DC-DC Regulator Module	From 13-16.8 VDC to 12 VDC used to power valves and NI module
DC-DC Regulator Module	From 6-24 VDC to 3-5-12 VDC used to power instrumentation
Pressure/Temperature Sensor	From 0 to 500 psiG and -40 to 125 C. Gems Sensors 3101H500PG02F001
Pressure/Temperature Sensor	From 0 to 1500 psiG and -40 to 125 C. Gems Sensors 3101H15CPG02F001
Pressure Regulator	Power Tank Pro Series regulator
Annular Flow Valve	GC Valves H401GF15Z1BF5
Secondary Flow Valves	Gems Sensors cryogenic valve B2011-LCO2
Gumstix Overo Fire COM	OMAP 3530 Applications Processor with wireless communications
Tobi Expansion Board	Expansion board for the Gumstix processor
NI WLS-9205	32-Ch, 16-Bit wireless voltage input for all sensor input

Figure 16 shows the piping and instrumentation of the propellant flow system for the prototype satellite. Caps of four CO₂ tanks were disassembled and machined to increase the inner diameter to achieve the required mass flow rates for the aerospike. The tanks are plumbed to converge into a manifold and up to a system regulator dropping the pressure to around 190 psiG. After pressure regulation, the tubing is split three ways to an annular flow valve and two secondary injection valves that are directly connected to the aerospike. Pulsing these valves control rotational movement of the satellite.

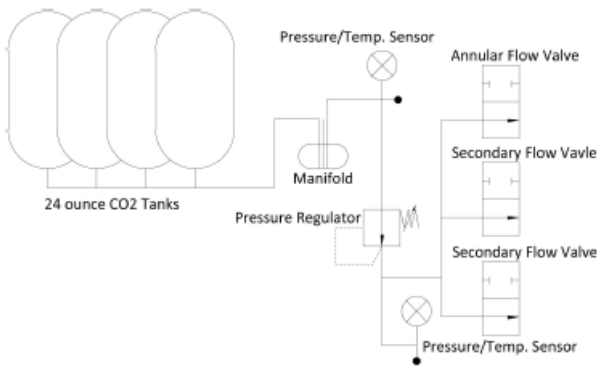


Figure 16. Piping and instrumentation of the propellant flow system for the satellite system.

C. Housing Design

Figure 17 shows the housing cage that was designed to suspend the satellite while keeping all coordinates fixed except for yaw. The

outer $2' \times 2' \times 4'$ frame was designed with 1" aluminum T-slotted framing with top and bottom supports and leveling feet. The housing cage encompasses the satellite leaving a gap of at least 6" on all sides of the satellite for maneuvering. Between the top of the satellite and the test cage, a gap of 15" was necessary to allow undisturbed annular flow for main plenum flow testing. A professional speedbag swivel from Everlast was attached to the top of the housing cage with a piano wire strung to hold the weight of the satellite while allowing it to rotate freely. This design allows the pitch and roll coordinates to be locked while mitigating the friction about the yaw axis to better simulate space conditions.

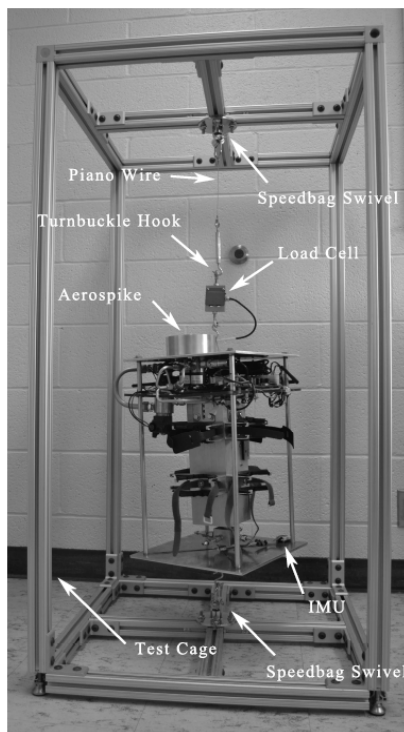


Figure 17. Satellite test stand.

D. Instrumentation

The satellite measures rotational angles, pressure, temperature, and various forces. Forces are measured at the manifold, pressure regu-

lator, and plenum. A National Instruments wireless 9205 data acquisition module is used to transmit measurements to a LabVIEW graphical user interface located on an off-satellite computer. The LabVIEW GUI is used for monitoring the system as well as for precheck tests. An on-board computer, Overo Fire Gumstix, holds the control laws and is directly interfaced to an inertial measurement unit and Gems Sensors cryogenic valves to provide a fast response time. Table III lists the parts with a further description of each major part.

The primary onboard navigation instrument is a miniature inertial measurement unit (IMU) built by Micro-Strain, Inc. [31]. The IMU features a high-performance attitude heading reference system that includes embedded tri-axial accelerometers, axis rate gyros, axis magnetometers, and a temperature sensor. The form factor and weight are very small, and this device is mounted on the inner platform of the vehicle without significantly affecting the weight and inertia of the platform. The IMU sensor data is blended internally, running a 'sensor fusion algorithm' to provide filtered data. User-selectable output parameters include Euler angles, direction cosine matrix components, acceleration vector components, three-axis angular rates, and three-axis magnetic field components. An interface circuit board was designed for the communication link between the IMU and Gumstix along with holding relays to actuate control valves. Figure 18 shows the block diagram for this communication. There is a two-way communication link between the Gumstix and the IMU to ensure correct data packets are being sent by comparing packet headings and checksums. Figure 19 shows the schematic diagram of the board, and Figure 20 shows the layout with the circuitry implemented.

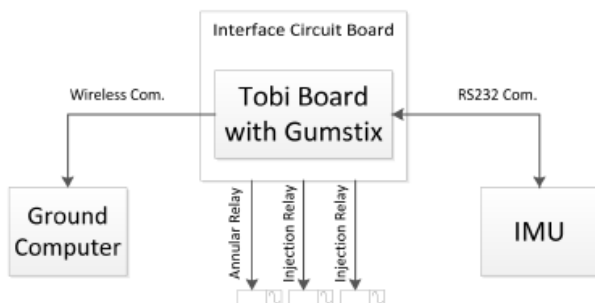


Figure 18. Block diagram of the communication link between the Gumstix, IMU, ground station computer, and relay blocks.

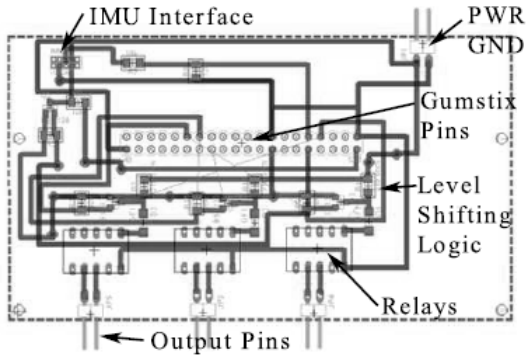


Figure 19. Interface circuit design for establishing a communication link between the IMU and the Gumstix along with relay outputs for solenoid valves.

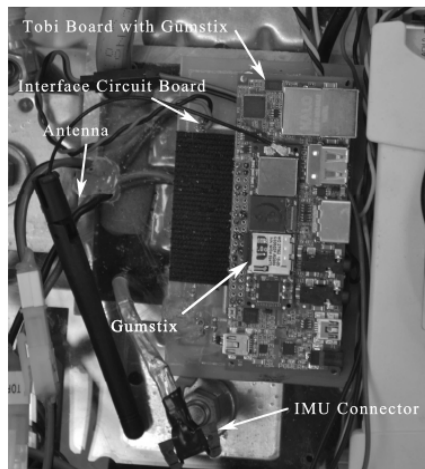


Figure 20. Interface circuit board with Gumstix, Tobi expansion board, wi-fi antenna, and IMU connector.

VII. Thrust Vectoring Tests

Both open- and closed-loop thrust vectoring tests were performed. The open-loop tests were used to quantify the delivered side force and moment levels and to work out system bugs. After the initial open-loop tests, rotational inertias of the system were measured, and a closed-loop

PID control law was developed and coded on the Gumstix avionics computer. Follow-on closed-loop control tests were used to adjust control gains.

A. Open-Loop Thrust Vectoring Tests

Initial tests consisted of placing the satellite at a resting state and turning on a single secondary injection port, causing the satellite to spin up to a desired angular rate of 2 rad/s (114.6 deg/s). Annular flow was included in the second test to compare the achieved amplification. These initial two test sets demonstrated the potential of aerospike vectoring to de-tumble artificial satellites. The third set of tests was used to ascertain a desired angle for satellite maneuvering. These tests were conducted with and without the presence of annular flow. Each ensemble of tests included at least eight tests to allow statistical verification of system performance and repeatability.

1) Open-Loop Tests with Secondary Injection Only

Each secondary injection test started at a fixed angular position with a rotational rate of 0 rad/s and ramped up until an angular measurement of 2 rad/s (114.6 deg/s) was read by the IMU. Once the desired rotational rate was reached, the test automatically turned off all valves and the data were saved in a packaged data set. A script was run to unpack the data into a format that could be imported into MatLab. Figure 21 shows the angular rate of the satellite compared with the time duration of each test. The secondary injection port produced a mean force of 4.19 newtons.

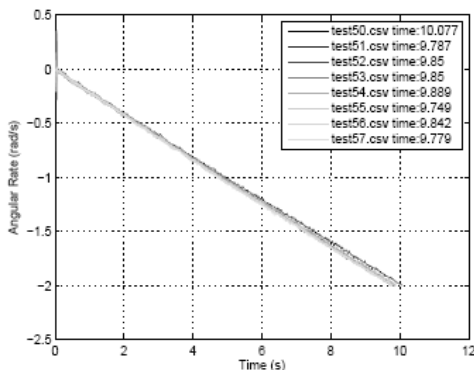


Figure 21. Spin-up tests conducted using only secondary injection for satellite acceleration from 0 rad/s to 2 rad/s.

The pressure and temperature in the plenum are at ambient conditions as the annular flow is not being used during these tests. All other pressures and temperatures varied slightly between tests but had no significant effect on the angular rate of the satellite. The time required to achieve 2 rad/s varied by 3.2% between the eight conducted tests.

2) Open-loop Tests Annular Flow with Secondary Injection

Each annular flow test was conducted like the secondary injection test with the exception of the usage annular flow. As the test was initiated, the annular flow was turned on while employing secondary injection until a rate of 2 rad/s was read by the IMU. Figure 22 displays the angular rate time history for each of the tests performed. With the annular flow active, the test results were moderately noisy. The filtered slope of each test was averaged to give an ensemble mean response curve. Figure 23 shows this result. The mean time to achieve a 2 rad/s rotational rate was approximately 15 seconds. This time difference was due to the swivel friction and off-axis swaying of the satellite produced by annular flow.

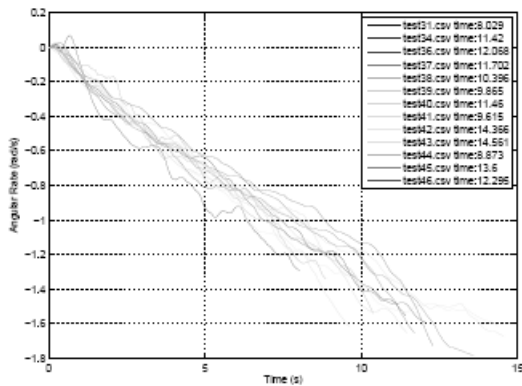


Figure 22. Spin-up tests with annular flow active and secondary injection accelerating from 0 rad/s to 2 rad/s.

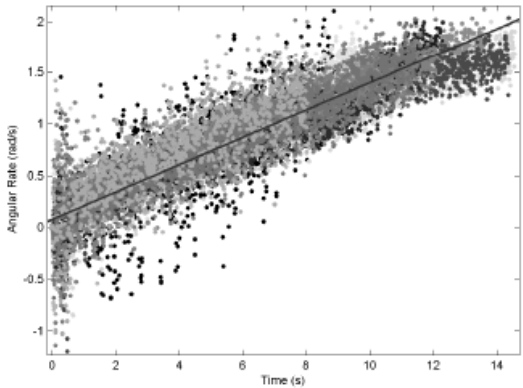


Figure 23. Angular rates for each data set with an averaged angular rate of 2 rad/s.

During this test series, the main plenum pressure and temperature varied considerably from test to test. No correlation between pressure and temperature was observed. Figure 24 shows this result. The observed time differences were primarily due to the swivel friction and off-axis swaying of the satellite. The aerospike’s location off the center axis of rotation caused a swaying motion during the tests. Figure 25 shows this off-axis motion.

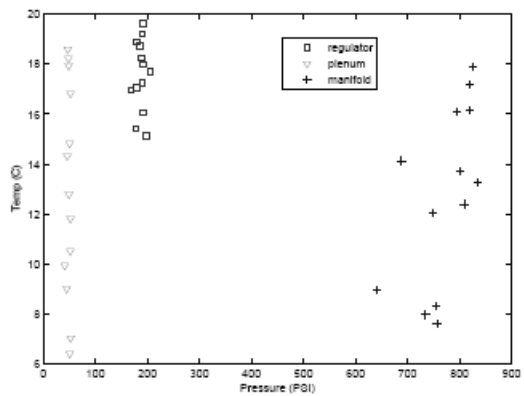


Figure 24. Pressure vs. temperature data for tests conducted using annular flow with a secondary injection for the regulator, plenum, and manifold.

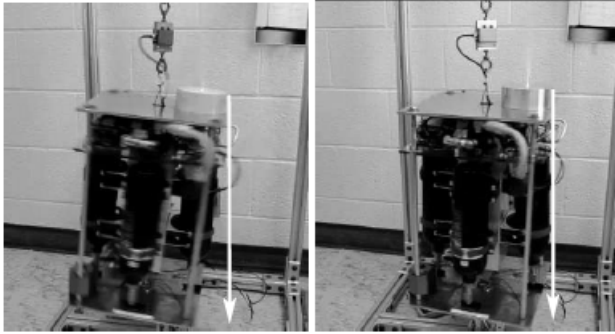


Figure 25. Comparison of the swaying motion produced by annular flow. (a) Sway that occurred near the end of a test. (b) Sway during the beginning of a test.

B. Closed-Loop Thrust Vectoring Tests

Closed-loop thrust vectoring tests consisted of placing the satellite at a resting state and controlling the secondary injection ports to move the satellite to a desired angular position. Tests conducted without annular flow ran for an 80-second duration whereas the tests conducted with annular flow ran for 35 seconds because of propellant constraints. These tests demonstrated the ability to vector the satellite in any direction with or without the presence of annular flow. Before any tests could occur it was important to understand the system dynamics in order to design good control laws for the system being controlled. An analysis of the system is provided in the below sections.

1) Moment of Inertial Measurements

It is essential to determine the natural frequency and damping ratio of the satellite before control theory is applied to the system. These constants were found by using a strong extension spring as a torsional spring. Figure 26 shows the 48.5-lb. satellite being suspended by the extension spring and rotated by weights on a pulley. The first test was used to calculate the spring torsion constant. Eight tests were conducted each placing a series of 10–100 grams of weight onto the pulley system while measuring the displacing of satellite from its resting position. By the displacement, the spring torsion coefficient was found by using Equation 1

$$K = \frac{-\tau}{\theta}, \text{ where } \tau = r * F \quad (1)$$

where τ = torque, θ = angle, r = radius, and F = force.

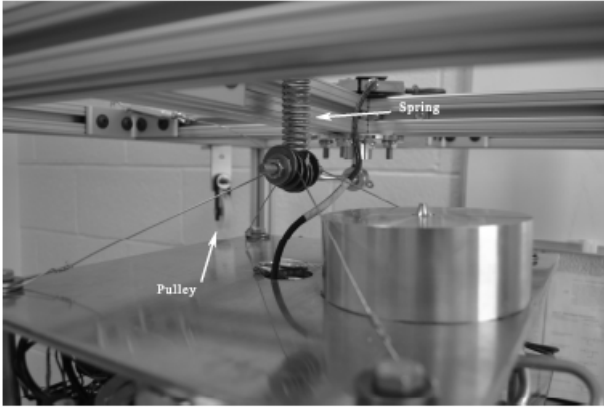


Figure 26. Test hardware to find the natural frequency and damping ratio of the satellite.

The natural frequency of the system was found by placing 100 grams on the pulley moving the satellite back to an angle of around 15 degrees. Once the satellite was at rest, the weight was quickly removed and the satellite was allowed to oscillate for 300 seconds. Angular position data was gathered and plotted in Figure 27a. The Fast Fourier Transform (FFT) of the data was taken to find the natural frequency of the system as shown in Figure 27b. Once the natural frequency was observed the damping ratio was found by implementing an exponential line fit and solving for ζ in Equation 2. I_{xx} was also found by solving Equation 3. Results were compared to and agreed within 3% of the moments of inertia calculated by Solid Edge.

$$y = X \exp(-\zeta \omega t) \quad (2)$$

$$\omega = \sqrt{\frac{k}{I_{xx}}} \quad (3)$$

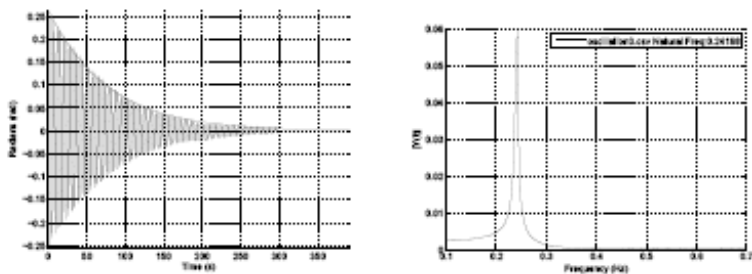


Figure 27. Test data collected for finding the natural frequency and damping ratio of the satellite. (a) Average oscillation of the satellite with the exponential line fit. (b) FFT of the average oscillation plot.

2) Secondary Injection Port Hardware Modifications

The initial machining of the aerospike was not completed according to specification. The base hole of the secondary injection port did not line up properly with the intersecting drill hole leading to the contour of the spike, resulting in almost no side force. To resolve this issue, a bigger drill bit was used to bore out the base hole wide enough to allow ample flow to the intersecting drill hole. With this modification, the open-loop tests were performed again, and 2 rad/s response times were approximately 1 second faster than the previously observed response times (Figure 28).

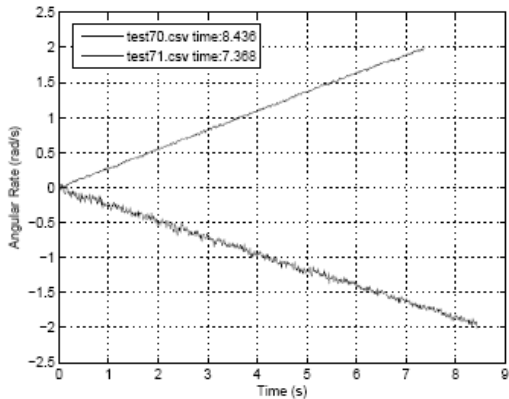


Figure 28. Test to compare the secondary injection port #1 with the secondary injection port #2. Port #1 was re-drilled to allow a greater mass flow to inject onto the contour of the aerospike.

3) Closed-Loop Control Law Development

The control law was developed using a commonly used feedback controller known as a PID controller. Figure 29 shows a block diagram of this controller. The PID controller is used to calculate and minimize the error by taking the difference between the actual and the desired process and adjusts the control output accordingly. For the satellite system, the desired process to control is the angular position. The feedback loop used the actual angular position from the IMU sensors and compared it with the desired position to calculate the total angular error. This error will be multiplied by each section of the PID controller and summed to produce the control output for the secondary injection ports.

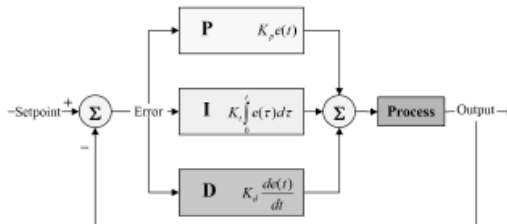


Figure 29. Block diagram of a typical PID controller [32].

Before control laws are applied, the system must be modeled. The satellite system is modeled by the relationship between desired angle and output torque. Equation 4 defines the system's open-loop transfer function where $0 \leq \zeta \leq 1$ where K is the system gain, w_n is the system's natural frequency, and ζ is the system's damping ratio.

$$G(s) = K \frac{w_n^2}{s^2 + 2\zeta w_n s + w_n^2} \quad (4)$$

A PID controller transfer function is defined in Equation 5, where K_p is the controller's proportional gain, K_i is the controller's integral gain, and K_d is the controller's derivative gain. Equation 6 is the systems characteristic equation.

$$Gc(s) = \frac{K_d s^2 + K_p s + K_i}{s} \quad (5)$$

$$q(s) = \alpha s^3 + (2\zeta w_n \alpha + K_d \alpha w_n^2) s^2 + (w_n^2 + K_p \alpha w_n^2) s + w_n^2 K_i \quad (6)$$

Control algorithms were developed and simulated using Mathsoft's MATLAB. Figure 30 shows a root locus plot of the system and plant while Figure 31 shows the controlled position using the PID controller described above. After developing the controller in MATLAB, the algorithms were implemented and processed on the Overo Gumstix via the interface board. After running through the PID controller, the control output is converted to a pulse-width modulation signal for the system output.

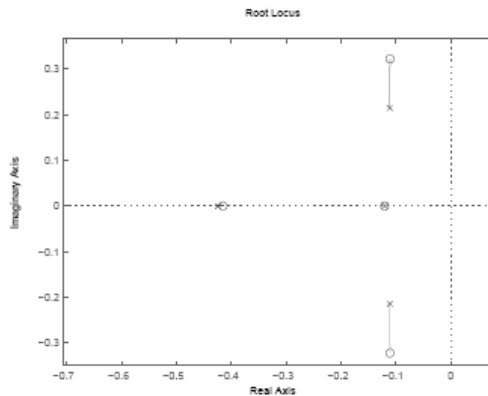


Figure 30. Root locus plot of the PID controller and plant. x represents the poles; o represents the zeros. The controller can be selected by selecting a point along the root locus that coincides with damping ratio and natural frequency.

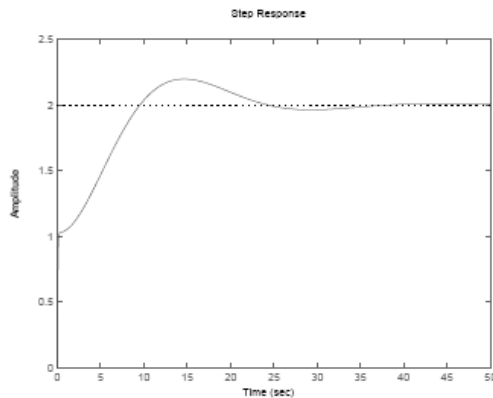


Figure 31. Controlled position using the PID controller tested in MATLAB.

4) Control Algorithm Implementation: An Overo Fire

Gumstix was used as the onboard computer for implementing the control algorithm. Upon this sequence the communication between the IMU sensor and Gumstix is established along with initialization of the PID control gains and control input. The control input is defined in rad as the angular position of the satellite. After initializing the system, the process steps into the control feedback loop. First, the control feedback loop requests information from the IMU sensor to describe the satellite's current angular position and angular rates. After this request, the information is read from the IMU and stored into buffers for computing. The current angular position is compared with the desired angular position, resulting in a total angular error. The PID controller computes the proportional, integral, and derivative errors separately. The proportional controller multiplies the error by the proportional gain. The integral controller multiplies total tests accumulated angular error by the integral gain. The derivative controller multiplies the angular velocity by the derivative gain. The PID calculations are summed to produce the controller output for the satellite system shown in Equation 7.

$$mb = K_d e(t) + K_i \int e(t) + K_d \frac{de(t)}{dt} \quad (7)$$

Because the valve is either active or inactive, a PWM conversion is implemented. This conversion is implemented by converting the PID control output to a set pulse length within the operating frequency range. The control feedback loop runs at 2 Hz. This frequency is set by the B-Series Cryogenic valves, which have an opening response time of 50 ms, thus the shortest pulse achievable for the valve to fully open and close is 0.1 seconds. The frequency is set to 2 Hz to allow the longest pulse to be five times the shortest pulse length. The PID control output becomes a factor of how long the valve is open during the 0.5-second cycle interval. The conversion is defined in Equation 8 and shown in Figure 32. The test runs for a predetermined time. After the time is reached all valves and communication ports are closed. The code was developed in python and located in the appendixes. Equation 8 limits commanded moments to positive values. Negative moment commands are generated by firing the opposing vectoring port.

$$V_o \begin{cases} \text{if} & mb \leq 0 \\ \text{elseif} & 0 < mb < max \\ \text{elseif} & mb \geq max \\ \text{else} & \end{cases} \begin{cases} V_{o=0} \\ V_o = mb \\ V_o = max \\ V_o = 0 \end{cases} \quad (8)$$

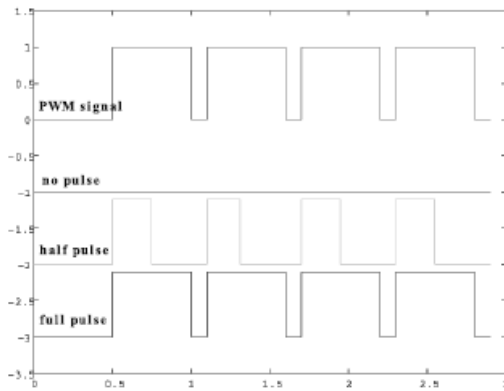


Figure 32. PWM signal shaping for (a) no pulse, (b) half pulse, and (c) full pulse.

5) Simulated Closed-Loop Positioning Results

Figures 33, 34, and 35 show results from MATLAB simulations using a PWM controller. The satellite can be controlled with or without the presence of annular flow.

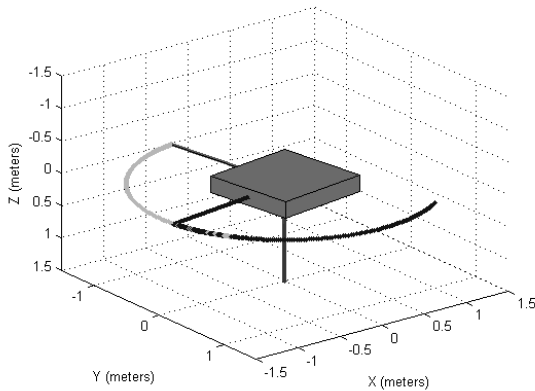


Figure 33. MATLAB simulation of graphical output using a PID controller.

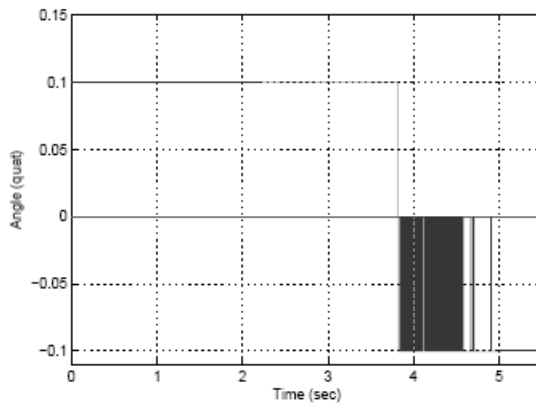


Figure 34. MATLAB simulation pulses sent to solenoid valves after PWM conversion.

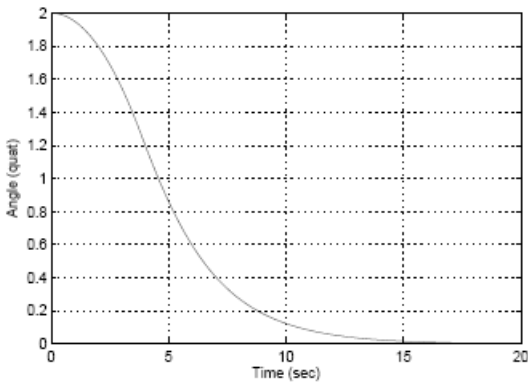


Figure 35. MatLab simulation angular error.

6) Closed-Loop Tests with Secondary Injection Only

Closed-loop vectoring tests with secondary injection were conducted using both a PI and PID controller. Damping was initially neglected. Figure 36 shows the PI controller in an oscillatory state, due to the lack of appreciable damping. As the derivative term was added into the controller, the plant's oscillation significantly reduced (Figure 37). The initial gains were derived from the natural frequency of the system,

but later manually tweaked to provide better control. The gains for the PID controller are indicated in Table IV. The tests started around 0 rad with a control input angle of 1.56 rad. The tests were very reliable and repeatable. All pressures and temperature were consistent throughout the test duration.

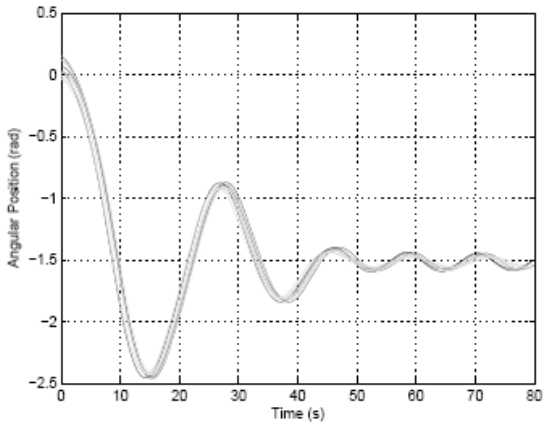


Figure 36. Position test using only the secondary injection ports with a PI controller.

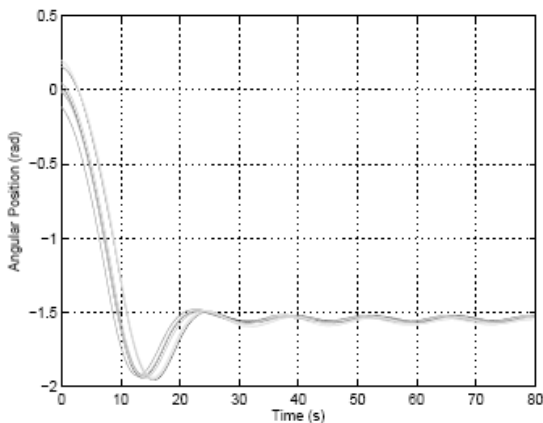


Figure 37. Position test using only the secondary injection ports with a PID controller.

Table IV
CONTROL GAINS

Control Gains	Value
Kp	5
Ki	1
Kd	-2.7

7) Closed-Loop Tests with Annular Flow and Secondary Injection

Closed-loop thrust vectoring tests with annular flow and secondary injection were performed using a PID controller with a control input of 1.56 rad. Initially, gains from the previous tests were used and the derivative gain was manually adjusted to give the desired response. This damping factor slowed down the rotation of the satellite when annular flow was on. Figure 38 shows the angular position of the PID controller vs. time. The angular rates were also plotted in Figure 39 to show when the secondary pulsing took place. Each sharp ridge indicates a pulse. A single angular rate was plotted in Figure 40 to better view the pulses of a single test. As time increases, or the test satellite moves closer to the desired position, the pulses become more rapid and multidirectional to hold the position. Figure 38 shows closed loop control using secondary injection ports with annular flow.

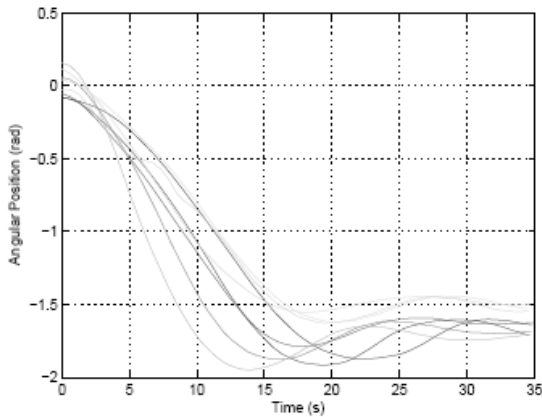


Figure 38. Position tests using annular flow and secondary injection with a PID controller.

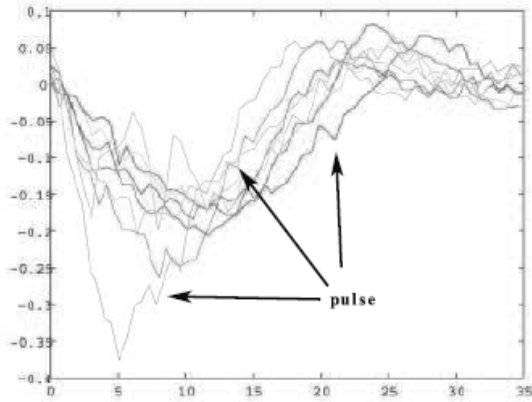


Figure 39. Angular rates during the position test using annular flow and secondary injection with a PID controller.

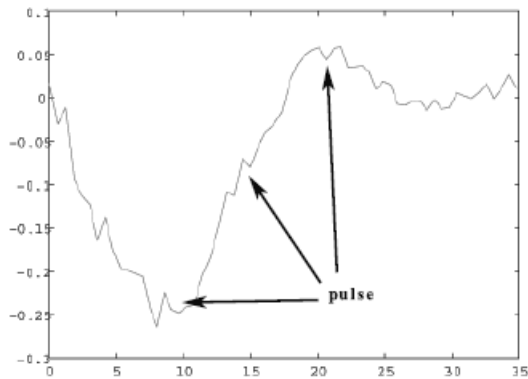


Figure 40. Angular rates from a single position test using annular flow and secondary injection with a PID controller.

VIII. Summary of Experimental Results

This article summarizes the experimental tests conducted during this research. Comparisons are made for secondary injection angular control with and without annular flow for open-loop and closed-loop tests.

A. Secondary Injection Comparison

Tests conducted without annular flow were very smooth and repeatable even with a slight variance in pressure and temperature. Predicted results concluded that the tests with annular flow would reach 2 rad/s faster than without annular flow. This, however, was not the case as tests with primary flow took an average of 5 seconds longer to reach the desired angular rate with the secondary injection port producing a mean force of 4.19 newtons. Figure 25 shows a swaying motion on the test rig during annular flow caused by the downward force produced by main plenum flow. Because the satellite hangs from the test stand, the off-centered aerospike causes the satellite to precess in the direction of the force exerted from the plenum. Because of this precession, a direct comparison cannot accurately be made. The tests proved the ability to rotate the satellite with and without annular flow. Pulsing from secondary injection port only can maneuver a satellite.

B. Experimental Positioning Results

Tests were conducted by releasing CO₂ through secondary injection ports moving the satellite from an initial position of 0 rad to a final position of 1.56 rad. Figure 37 shows that without annular flow the system is very reliable and repeatable, indicating that attitude control using an aerospike nozzle on a small satellite is feasible. Control positioning tests with annular flow are also repeatable and reliable within tolerance. These results are not as clean as the results produced without annular flow because of the swaying motion produced in the test stand. This inconsistent test apparatus artifact must be corrected when developing a full 3-degree of freedom (3-DoF) test article. Generally, the presented results demonstrate the feasibility of controlling a small satellite using secondary injection ports on an aerospike nozzle.

IX. Conclusion

This paper supports the concept of using secondary injection ports on an aerospike nozzle to control a prototype CubeSat. Thrust vectoring tests were performed using open-loop and closed-loop control. Open-loop control tests were used to quantify the delivered side fore and moment levels as well as correcting any system bugs. During these tests, a secondary injection port was used to prototype satellite spin up to an angular rate of 2 rad/s. Consistently achieved angular rate demonstrated high control fidelity and repeatability. The time required to achieve 2 rad/s varied by 0.328 seconds between the eight conducted tests with a mean time of 9.85 seconds. The same tests conducted with

annular flow were moderately noisy and demonstrated less control fidelity. The mean time to reach a angular rate of 2 rad/s was 15 seconds. The observed time differences between tests with and without annular flow were primarily due to the swivel friction and off-axis swaying of the satellite caused by annular flow. Closed-loop thrust vectoring tests consist of placing the satellite at a resting state and controlling the secondary injection ports to move the satellite to a desired known position.

A PID controller was used for closed-loop vectoring control. Tests without annular flow were conducted for 80 seconds each. It was observed that positioning control is repeatable and reliable.

Closed-loop thrust vectoring tests with annular flow were run for 35 seconds. These tests were moderately noisy and varied more than the tests without annular flow because of the swivel friction and off-axis swaying of the satellite. Nonetheless, the prototype system is controllable and repeatable with a closed-loop PID controller.

The hardware in the loop demonstration realized the capability to precisely position a satellite in a single axis coordinate frame with and without annular flow. The initial tests have verified objectives 1 and 2, showing that rotational ramping is achieved from the secondary injection ports simulating the ability to de-tumble or position a satellite. Final tests have verified objectives 3 and 4, proving the ability to precisely position a satellite for a one-axis attitude control simulation. The above tests indicate that using an aerospike nozzle with secondary injection ports for attitude control is feasible.

A. Future Work

Immediate future work includes developing different control algorithms with a comparison test to see which provides the best desired control. More extensive future work for developing the aerospike nozzle to achieve a sufficient TRL level for flight is described below.

- Developing a prototype satellite integrating two aerospike nozzles for controlling one-axis of rotation.
- Closed-loop control demonstration using the aerospike nozzles for a 3-DoF simulation using cold gas.
- Hanging the prototype to a tethered balloon showing 3-DoF control using cold gas.
- Developing a scaled prototype satellite integrating two aerospike nozzles using hot gas.
- Testing control algorithms for a 3-DoF ground demonstrations.
- Development of a satellite system to fit within a 2U CubeSat.
- Testing control algorithms for a 3-DoF ground demonstration.

- Designing actual flight hardware for a CubeSat propulsion unit.

X. Recommendations

Below is a list of recommendations for those pursuing future research and development in this area of study. The list contains areas of the project that can be improved as well as lessons learned from the development process.

- Use high-pressure air instead of CO₂. This would elevate any freezing in the system and two-phase flow.
- Develop a better method for frictionless rotation. For example, consider developing an air table to suspend the satellite.
- If the propellant tanks are going to be taken on and off the system for refilling, consider using a stronger plumbing material than aluminum. Some of the piping was replaced several times because of the continuous stress of refilling propellant bottles.
- If running future tests with only one aerospike nozzle, consider possible solutions to mitigate the precession in the system. This would allow for a comparison to be made between the tests with annular flow and without.

Acknowledgment

I would like to thank Dr. Doran Baker and Dr. Stephen Whitmore and Dr. Donald Cripps for dedicating their time and support to further the research of this project, along with the Rocky Mountain NASA Space Grant Consortium and NASA's Office of Education through their EPSCoR funds they contributed to make this project possible. Special thanks go to all the team members I worked with throughout this project, Stewart Hansen, Shannon Eilers, and Nathan Madsen. I would also like to thank the machinists for the use of their equipment, machine time, and insights.

References

- [1] "Final report, advanced aerodynamic spike configurations, volume 1," Rocketdyne Advanced Projects, Tech. Rep., 1967.
- [2] "Final report—studies of improved Saturn V vehicles and intermediate payload vehicles," The Boeing Company—Space Division, Tech. Rep., 1996.
- [3] "Final report, advanced aerodynamic spike configurations, volume

2,” Rocketdyne Advanced Projects, Tech. Rep., 1967.

[4] C. Bendersky, Space Shuttle Propulsion Issue, Staged Combustion Bell Versus Tap-Off or Gas-Generator Aerospike, Bellcomm. Inc. Std.

[5] K. C. Hendershot, R. J. Sergeant, and H. B. Wilson, “A new approach for evaluating the performance and base environment characteristics of nonconventional rocket propulsion systems,” AIAA, vol. AIAA-67-256, 1967.

[6] J. J. Korte, A. O. Salas, H. J. Dunn, N. M. Alexandrov, W. W. Follett, G. E. Orient, and A. H. Hadid, “Multidisciplinary approach to aerospike nozzle design,” National Aeronautics and Space Administration, Langley Research Center, Tech. Rep., 1997.

[7] K. Higdon and D. B. Landrum, “Analysis of annular plug nozzle performance and tvc,” AIAA paper, vol. AIAA-2003-4908, 2003.

[8] H. F. R. Schoyer, “Thrust vector control for (clustered modules) plug nozzles: Some considerations,” Journal of Propulsion and Power, vol. 16, pp. 196–201, 2000.

[9] J. Ruf and D. McDaniels, “Experimental results for an annular aerospike with differential throttling,” in 5th International Symposium on Liquid Space Propulsion, 2003.

[10] M. Fick and R. H. Schmucker, “Remarks on plug cluster nozzles,” in 31st AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, vol. AIAA 95-2694, 1995.

[11] G. Hagemann, C.-A. Schley, E. Odintsov, and A. Sobatchkine, “Nozzle flowfield analyses with particular regard to 3d-plug cluster configurations,” AIAA, vol. AIAA-1996-2954, 1996.

[12] R. Sorge, C. Carmicino, and A. Nocito, “Design of a lab-scale cooled two-dimensional plug nozzle for experimental tests,” 38th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, 2002.

[13] F. Nasuti, M. Geron, and R. Paciorri, “Three dimensional features of clustered plug nozzle flows,” in 39th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, 2003.

[14] T. Ito, K. Fujii, and G. Hagemann, "Numerical investigation of the sideforce effect on linear plug nozzle performance," AIAA, vol. AIAA 2004- 4018, p. all, 2004.

[15] T. Ito, K. Fujii, and A. Hayashi, "Computations of the axisymmetric plug nozzle flow fields: Flow structures and thrust performance," in 17th AIAA Applied Aerodynamics Conference, no. AIAA 1999-3211, 1999.

[16] H. Miyamoto, A. Matsuo, and T. Kojima, "Effects of sidewall configurations on rectangular plug nozzle performance," AIAA, vol. AIAA-2006- 4373, p. all, 2006.

[17] T. Ito and K. Fujii, "Flow field and performance analysis of an annular type aerospike nozzle with base bleeding," Transactions of the Japan Society for Aeronautical and Space Sciences, vol. 46-151, pp. 17–23, 2003.

[18] —, "Numerical analysis of the base bleed effect on the aerospike nozzles," Transactions of the Japan Society for Aeronautical and Space Sciences, vol. 46-151, pp. 17–23, 2003.

[19] —, "Flow field analysis of the base region of axisymmetric aerospike nozzles," in 39th AIAA Aerospace Sciences Meeting & Exhibit, no. AIAA 2001-1051, 2001.

[20] A. N. Kraiko and N. I. Tillyayeva, "Contouring spike nozzles and determining the optimal direction of their primary flows," Fluid Dynamics, vol. 42-2, pp. 321–329, 2007.

[21] N. Karthikeyan, S. B. Verma, and L. V. Venkatakrishnan, "Experimental investigation of the acoustics of an annular aerospike nozzle flow," in 15th AIAA/CEAS Aeroacoustics Conference (30th AIAA Aeroacoustics Conference), 2009.

[22] S. C. Shark, J. D. Dennis, and J. K. Villarreal, "Experimental performance analysis of a toroidal aerospike nozzle integrated with a n2o/htpb hybrid rocket motor," in 46th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, 2010.

[23] J. R. Stoffel, "Experimental and theoretical investigation of aerospike nozzles in a hybrid rocket propulsion system," in 47th AIAA Aerospace Sciences Meeting, 2009.

- [24] P. Lemieux, “Nitrous oxide cooling in hybrid rocket nozzles,” *Progress in Aerospace Sciences*, vol. 46, pp. 106–115, 2009.
- [25] E. Besnard and J. Garvey, “Aerospike engines for nanosat and small launch vehicles (nlv/slv),” in *Space 2004 Conference & Exhibit*, 2004.
- [26] —, “Development and flight-testing of liquid propellant aerospike engines,” in *40th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit*, 2004.
- [27] S. D. Eilers, M. D. Wilson, Z. W. Peterson, and S.A. Whitmore, “Analytical and experimental evaluation of aerodynamic thrust vectoring on an aerospike nozzle,” *46th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit*, Nashville, TN, 2011.
- [28] A. Naghib-Lahouti and E. Tolouei, “Investigation of the effect of base bleed on thrust performance of a truncated aerospike nozzle in off-design conditions,” in *European Conference on Computational Fluid Dynamics*, 2006.
- [29] W. A. Jordi Puig-Suari, Clark Turner, “Development of the standard CubeSat deployer and a CubeSat class picosatellite,” in *Aerospace Conference*. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=931726>: IEEE, 2001.
- [30] T. Angstadt, “Fortran program for plug nozzle design,” 2001, 2001; A written assignment for a course by the name TECH 591.
- [31] M. Inc., “3dm-gx3-25, miniature attitude, heading reference system (ahrs),” June 2010. [Online]. Available: <http://www.microstrain.com/3dm-gx3-25.aspx>
- [32] SilverStar. (2006, October) Depicts a traditional pid controller. [Online]. Available: <http://en.wikipedia.org/wiki/File:Pid-feedback-nct-int-correct.png>

BYU's Entry in the University Rover Challenge: Making a Simulation Mars Rover

Alexander R. Jafek, Andrew K. Norgrant, and Justin S. Smith*

Brigham Young University

Abstract

The University Rover Challenge is an annual competition held by the Mars Society in Hanksville, Utah. The purpose of the competition is to promote the design of prototype Mars Rovers whose design creativity and implementation will serve as a proof-of-concept for future rovers intended to assist humans in their explorations of other planets. Brigham Young University (BYU) has sent a rover to compete each of the last four years. This paper will present a brief history of the project as well as our work in solving the many mechanical and electrical engineering challenges we have faced. We address the mechanical aspects, including the chassis, suspension, and arm designs. We also outline our control, communication, and live video/sensory feedback systems. Based on our research, we have designed and will present a practical

* All authors contributed equally.

rover ready to compete in this year's University Rover Challenge with the desire to progress our knowledge of feasible design concepts to be implemented in future Mars rovers.

I. Introduction

The University Rover Challenge is an annual competition hosted by the Mars Society in Hanksville, Utah. Each year since 2007, the Mars Society has challenged universities to design, build, and compete with robotic rovers. These rovers serve as proof-of-concept designs for "the next generation of Mars Rovers that will one day work alongside human explorers in the field."¹ The competition requires that rovers not exceed 50 kilograms in mass nor have a net value of over \$15,000. To simulate Martian terrain, the competition is held during the summer in the Utah desert, where temperatures surpass 100° F.

In the competition, we will face four distinct challenges using our rover as a stand-alone platform, controlled from a command station without visual contact. The four tasks, provided directly from the Mars Society, are:

1. Site Survey Task

"Teams shall be required to perform a remote survey to determine the precise coordinates of field markers that are not reachable by the rover."¹

2. Sample Return Task

Teams are to "collect and return a single sample from the site they determine to have the greatest likelihood of containing photosynthetic bacteria"¹ or other signs of life.

3. Astronaut Assistance Task

"Teams shall be required to deliver multiple supply containers to simulated astronauts in the field as quickly as possible."¹

4. Equipment Servicing Task

"Rovers shall be required to perform several dexterous operations on a mock-up equipment panel."¹

The competition has grown to attract teams from around the world. BYU has competed every year since the competition's establishment. This has given us the opportunity to discover what works and

what doesn't in the competition's extreme conditions. In this paper, we present some of the innovations we have made to overcome the mechanical and electrical engineering challenges we have faced. We present first our mechanical observations, including our development of a robust chassis and suspension system as well as comments on our development of a robotic arm. We then present the development of our control and communication/video systems. Although we recognize that our design is constantly changing, we provide our most accurate and current research results. We believe our research has the potential to be integrated into future interplanetary explorers.

II. Chassis and Suspension

A. Chassis Requirements

Our rover will be required to traverse over simulated Mars landscape, including soft terrain, varying debris, and 50% slopes.

B. Previous Work and Observations

1. Problems of a Rigid Chassis

BYU's 2007 Mars rover used a rigid chassis. In testing our rover, and in observing others, we have found that this hindered performance because it minimized contact with the ground as the chassis could not contour to the terrain.

2. Narrow Wheels Cause Problems

We have noted that rovers with narrow wheels frequently get stuck in the loose dirt of the simulated Martian environment.

3. Electronics Need to be Accessible

Problems with electronics on the day of the competition have plagued BYU, particularly in 2008, and many other teams. Consequently, it is necessary to have easy access to the electronics on the rover.

4. Disadvantages of Complex Drive Linkage Systems

Although BYU has avoided complex drive linkage systems from the beginning, their shortcomings have been demonstrated by competitors' failed attempts to employ chains and bearings.

5. High Ground Clearance

A high ground clearance improves all-terrain performance and allows us to worry less about obstacles.

6. Low Center of Gravity

A low center of gravity is essential to maintaining stability. For this reason, we have striven to balance this requirement while maintaining a high ground clearance.

7. The Advantages of Beach Wheelchair Tires

We have observed the many advantages of beach wheelchair tires. First, the tires provide a large amount of ground-contacting surface area, thus increasing stability. Second, the tires absorb the minor vibrations that can be problematic for the electronic systems. Third, the tires are soft enough to allow our skid steering system to work effectively.

8. Problems of Planetary Transmissions

Although they are easy to design, we have found that planetary transmissions are particularly vulnerable to the harsh conditions of the test site. At last year's competition, seven transmissions were severely damaged as planetary gearboxes loosened or had teeth sheared off their interior gears.

B. Concept Generation

Having established our general requirements, we began to formulate chassis designs that could fulfill all of our needs. Four major suspension concepts were researched and evaluated based on the design specifications: four wheeled, treaded, rocker-bogie, and six wheeled. Below is a brief overview of the thought processes behind each design.

1. Four Wheeled

Our previous rover design consisted of four A-arm type suspension arms with planetary gearboxes around a rectangular chassis as pictured in Figure 1. It was a robust and effective design. Nevertheless, it does not leave much room for innovation.



Figure 1. BYU's competition chassis 2009.

2. Treaded

A computer assisted design (CAD) of the proposed treaded design is presented in Figure 2. The major advantages of the treaded, tank-like design are the traction, learning factor, and aesthetic appeal. Some of the disadvantages of this design are the complexity, potential lack of reliability in the treads, maneuverability, weight, and clearance. These factors combined to eliminate the novel and challenging idea of a treaded rover.

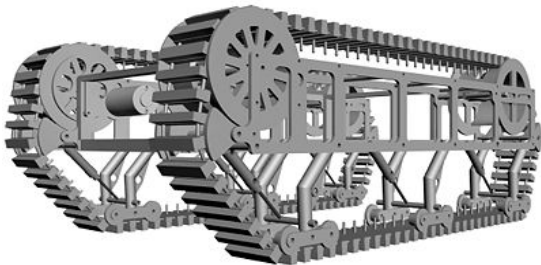


Figure 2. CAD model of proposed treaded design.

3. Rocker-Bogie

The rocker-bogie suspension is used by NASA² for the current Mars Rovers and is pictured in Figure 3. Its primary advantage is the articulation and range of motion of the individually powered wheels; this allows the suspension to climb an obstacle up to twice the diameter of the wheels in a stable manner. It was not, however, designed for high speeds or skid steering, conditions under which the rover will operate.

Damping suspension modifications could make this a very viable design for future rovers.

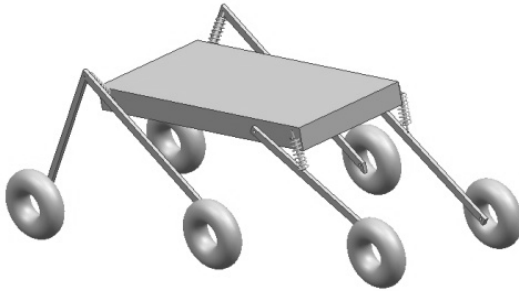


Figure 3. CAD model of proposed rocker-bogie design.

4. Six Wheeled

A six-wheeled suspension, shown in Figure 4, with a trailing arm was also proposed. Its six wheels would provide excellent traction, power, and stability. We found this design advantageous, although the added number of wheels adds weight and requires that we find replacements for our previous failing planetary gearboxes. Therefore, this basic design was selected on the premise that with a simpler design, more time could be spent engineering it to high performance and reliability rather than trying to design a lesser-known system that might not perform as well.



Figure 4. CAD model of proposed six-wheeled design.

To generalize our findings, we have determined that special protection needs to be provided for the motor and gear boxes because of

the dust and heat and that, although there are many ways to effectively traverse the terrain, it is best to optimize a practical and well-known design.

C. Detail Design

After selecting a basic design to elaborate on, we spent our time creating models, performing analysis, and choosing components aimed to meet the specifications established previously. We first selected the components that would be commercially purchased and then custom designed the rest of our chassis around these specifications. Since most of our rover was to be custom designed and machined at BYU, it was necessary to first get the dimensions of parts to be purchased. We determined that we could save money and obtain higher quality by simply buying the first two components: tires and motors.

1. Tires

As was mentioned in the previous discussion, we concluded that beach wheelchair tires would best serve our rover. After extensive market research, we settled on the Wheeliez 24cm PU Beach Wheel. Designed for operation in sand, these tires are well suited for the competition terrain. We therefore account in all of our future designs for the tires' specifications: 10-inch diameter and 1.5-pound weight.

2. Motors

In selecting our motors we optimized torque, gear ratio, and speed while minimizing weight and size. To do this, we determined the necessary torque, found a motor of that specification, did calculations to determine our gear ratio, and then ensured that this system would provide adequate speed. Commentary on this process, as well as calculations follow.

- i. Torque:* The competition requires that our rover be capable of climbing a 50% slope (roughly 23°). We concluded that we could save time in the tasks by ensuring that our rover exceeded the minimum requirements of this challenge. We did calculations so as to design a rover that could accelerate at 1 m/s^2 up a 45° slope. The total output torque required was calculated to be 61.5 Nm, or 1451 in-oz per wheel.
- ii. Gear Ratio:* We matched these specifications with the F-P 00801-0673 Motor. The gear ratio required was determined to be in the range of 290:1 to 20:1, which gave us the ability to work with a variety of gear ratios.

iii. *Speed:* Our goal was a 10-mph top speed. This would require that our wheels rotate at 350 revolutions per minute (rpm). Working backwards from the specifications of our motor, we determined to use a gear ratio of 64:1. Some calculations show that meeting our goal of 10 mph would require a motor speed of 22,000 rpm, slightly beyond our motor's capabilities. Nevertheless, our rover will be capable of traveling 9.3 mph on flat ground. The use of a common gear ratio makes replacing parts, doing general calculations, and making modifications easier.

3. Chassis Components

After determining the system's tires and motors, we began to design the rest of the chassis around these elements.

i. *Aluminum Cradle:* The aluminum cradle provides a mounting surface for the components of the rover, including the A-arm suspensions, electronics box, antennas, and the package deployment pieces. The aluminum cradle is built from ½-inch square tubing welded using the tungsten inert gas process. It provides the outline and structural integrity of our rover (Fig. 5). It establishes the basic dimensions of our rover as 36" × 24".

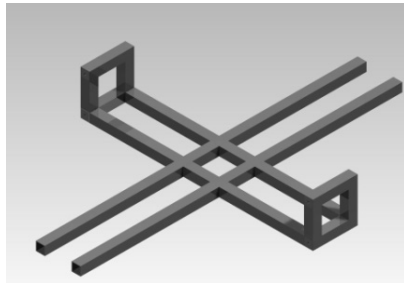


Figure 5. CAD model of aluminum cradle.

ii. *Aluminum Box:* The aluminum box provides the majority of the surface area on the top plane of our rover and has walls around the edge as shown in Figure 6. Our electronics will be placed on this flat sheet inside the rover. By putting all of the electronics in one central and accessible location, we make it simple for adjustments and repairs to be made.

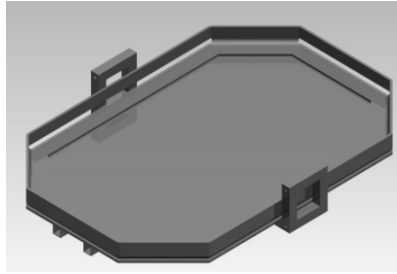


Figure 6. CAD model of aluminum cradle and aluminum box.

- iii. Cowl:* Mounted to the top of the box is an aluminum cowl as shown in Figure 7. The cowl provides a number of benefits for the electronics, namely keeping out dust, providing shade, and allowing cooling by a centralized system. We will also coat the cowl in a reflective material to protect the electronics from solar radiation.

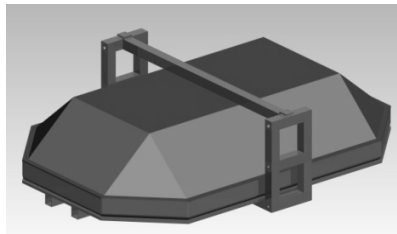


Figure 7. CAD model of aluminum cradle, box, and cowl.

4. Suspension

Suspension has proven to be one of the most critical selections in the chassis design, and as such, very careful considerations have been taken. The design features a central A-arm suspension and two compliant leaf springs suspensions, each attached to two other wheels. Figure 8 provides a detailed view of this design.

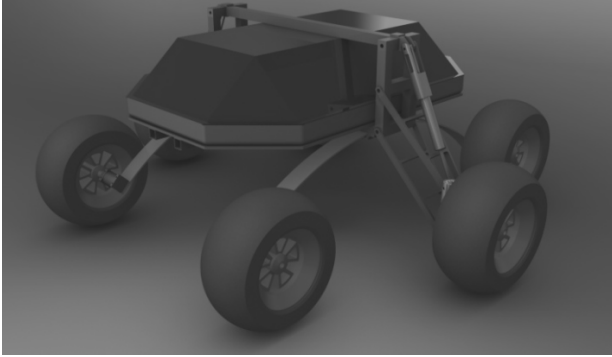


Figure 8. CAD of current rover design.

- i. A-arm:* A 30-pound gas spring, the Suspa C16-18974, provides the downward force, and the motor mounts directly to the arm, as shown in Figure 9. The arm components are cut from 0.25-inch 6061 aluminum using the BYU water jet.

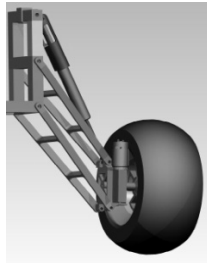


Figure 9. CAD of A-arm suspension.

- ii. Leaf Spring:* The other four wheels attach to a leaf spring-compliant mechanism (Fig. 10a), which is made from layered sheets of 3m 7781 E-Glass/SP381, a pre-impregnated fiberglass weave. The leaf spring attaches to the cradle using a clamp system as shown in Figure 10a. A thin piece of rubber will buffer the leaf spring from the cradle. Force element analysis (FEA) of the leaf spring concept proved its feasibility (Fig. 10b). Early testing with fiberglass proved that both 3- and 5-layer sheets were insufficient in providing the necessary upward force. Current tests are underway on a thicker leaf spring.

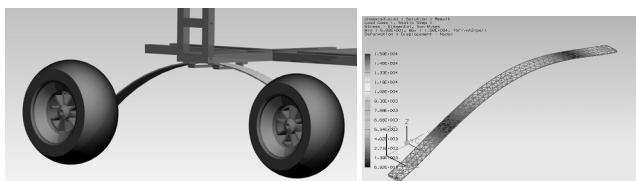


Figure 10. (a) CAD of leaf spring suspension. (b) FEA of leaf spring suspension. Darker areas represent areas of low stress, while the lighter areas have high stresses.

It should be noted that based on weight calculations in CAD systems, the entire chassis and suspension should be no more than 35 pounds; this is also well within specifications.

3. *Summary*

The six-wheeled design is superior to past rovers because of its additional traction and power. Using an A-arm for the middle wheels and a compliant leaf spring suspension for the other wheels is a system that we are confident will work in our rover and has great potential in the aerospace industry.

III. Arm

A. *Arm Requirements*

We need to construct an arm capable of pushing buttons, flipping switches, and connecting male 3-prong plugs (U.S. style) into vertical electrical outlets.

B. *Previous Work and Observations*

1. **Need for Custom-Made Arm**

In BYU's first entry at the University Rover Challenge in 2007, we used a commercially available arm. Although it performed well for us for that year, we determined that to perform the more complicated tasks of future years and to increase the size of the arm, we needed to custom-make our arm.

2. Need for a Versatile Gripper

Our gripper will be used to complete the service equipment servicing task, as well as to collect and return small rocks for the sample return task.

3. Problems of a Motor-Controlled Arm

In last year's competition, we competed with an arm whose movement was driven by motors. This resulted in a number of problems. First, the motors have no positional feedback system. Therefore, we had to develop an algorithm to determine the claw's position in three-dimensional space. Second, the motors had no limit of rotational motion. Unfortunately, this gave the motors the ability to rip the arm apart, as was demonstrated at the last competition.

C. Concept Generation

To provide an arm with the most flexibility in performing the competition tasks, we decided to build an arm with five degrees of freedom. We designed an arm with three major joint locations: shoulder, elbow, and wrist. The shoulder grants arm sweep and shoulder swivel, the elbow provides extension, and the wrist allows pitch and roll. All motions are controlled by servos, which natively provide feedback control for easy interfacing.

D. Detail Design

For reference we provide an image of the arm complete with the shoulder and elbow joints (Fig. 11).

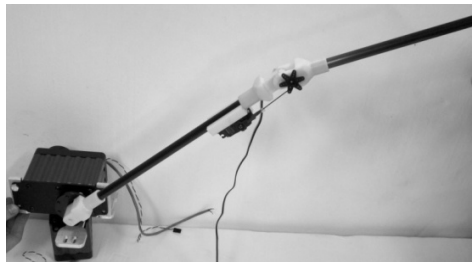


Figure 11. Full mechanical arm.

First, we determined the materials of the arm. The lengths of black tubing are carbon fiber, and the off-white joints are cut from ABS plastic. These materials are strong but lightweight.

After selecting the material for the arm, we began to work on the motion capabilities.

1. Shoulder

In order to get two degrees of motion from the shoulder, we stacked two servos on top of each other as shown in Figure 12. The bottom servo allows the arm to rotate relative to the top plane of the rover. The arm attaches directly to the top servo, which allows the angle of the entire arm to be changed relative to the top plane of the rover.

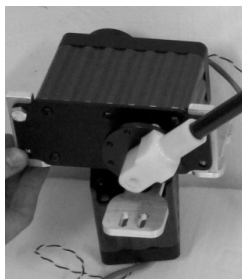


Figure 12. Shoulder joint.

Calculations proved that the large moment created by the long arm would mandate servos that could generate 1200 ounce-inches of torque in this joint. For this reason, the Invenscience i00600 Torxis Servo has been selected. It is fully capable of fulfilling our requirements as it can provide 1600 ounce-inches of torque.

2. Elbow

The elbow joint consists of both the joint itself and the system to move it. The joint, shown disassembled in Figure 13, is a pin joint with all the components, including the pin, made from ABS plastic. The system for extension, as shown in Figure 14, is a small servo mounted onto the lower part of the arm. This servo, the Futaba S3170G Digital Planetary Gear Retract Servo, provides 94 oz-in of torque and has proven to be very successful at its role in the system.

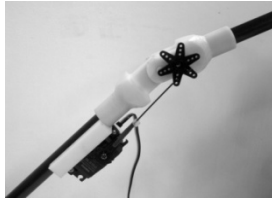


Figure 13. Elbow pieces.



Figure 14. Constructed elbow joint.

3. Wrist

The wrist is currently under development but when completed will provide both the roll and pitch for the gripper.

E. Summary

We have had great success in our shoulder and elbow joints that are quite light and rely on servos to move. The move away from motors to servos has improved the rover design, and the required adaptation of lighter materials has not been problematic.

IV. Control Systems

A. System Requirements

The purpose of the control system is to provide a reliable and intuitive interface between the operator and the rover's systems. This can be divided into two physical components: rover-side, which resides on the rover, and control-side, which resides at the base station. These two components must work together to execute commands and to provide feedback.

In executing commands, the control-side must take input from the operator and send it to the rover, while the rover-side must receive and act accordingly. In providing feedback, the rover-side must gather data and relay the information to the control-side, which then displays it to

the operator. In all aspects, the system must be reliable and implement fail-safes to prevent it from injuring itself or others.

B. Previous Work and Observations

1. Movements Away From Centralized Computing

The first consideration has been whether to implement the control system using a centralized computer or a modularized network of microcontrollers. In the past, every BYU competition rover has used a centralized computer, allowing for easier programming and greater processing ability. This centralized computer was a Windows machine in 2007, but it has been a Linux machine since.

Unfortunately, using a computer to run everything also meant that there was a single point of failure. In our case, the computer became extremely unreliable, likely as a result of overexposure to heat and power fluctuations. As a result, components were gradually transferred off the computer and onto a network of Arduino microcontrollers.

2. Motor Control Refinement

Control of the motors has been done through a Lynxmotion SSC-32 servo controller via Castle Creations Mamba Max speed controllers. This system works reliably in most situations. The main problem we have encountered is that the servo controller sends a command to the speed controller until it receives a new command. This becomes problematic if the rover drives out of communication range because it will continue to drive, and operators have no power to stop it. Thus, a control system must have a way to stop the motors if the rover goes out of range.

3. Some Motor Controllers Have Innate Problems

Another consideration for our new system is that some speed controllers switch to a programming mode if they receive a specific sequence of instructions.

4. Ability to Compensate for Human Error is Necessary

Other previous problems were more the result of poor actions by the operator. For example, sudden movements of the joystick have caused significant amounts of force to be exerted on the gear boxes, resulting in damage. The control system should take steps to prevent such damage from occurring.

5. Difficulties of Arm Control

Control of an arm has always been problematic. We had researched reverse kinematics but eventually concluded that the programming was simply too complicated given our time frame. As a result, we built a custom interface consisting of knobs to individually control each joint. The optimal solution would be a scale model of the arm with potentiometers at each joint. This would allow us to have the real arm mirror the model.

C. Concept Generation

1. Rover-Side Processing

The largest debate for rover-side processing was the use of an on-board computer versus the current setup of microcontrollers.

The primary advantage of microcontrollers is that they only run our code and as a result have an edge in reliability. Additionally, power consumption and heat generation are lower. The main problem is that there is little reliable code available to interface between the microcontrollers and rover components. Although this system has worked in the past, it required excessive amounts of time to learn and implement the protocols to send instructions over a network and to communicate with peripherals.

The main question regarding a computer was how it would interact with the rover. National Instrument's LabVIEW has been used for autonomous vehicle control (Dr. D. Wilde, personal communication).³ After a little investigation, we discovered that LabVIEW has built-in support for the servo controller that we have been using and for interfacing with networks, joysticks, a global positioning system (GPS), and other components. The graphic programming environment allowed us to utilize greater levels of abstraction in our code (see "Detailed Design" below). Additionally, a graphic user interface is automatically created as the code develops, providing an easy method to build the control interface.

Based on the discovery that LabVIEW is more fully featured on Windows, we switched to a computer running Windows XP Pro. The rover computer is based on a microATX motherboard running an Intel Core 2 Duo T7400 2.166-GHz processor with 2 GB of memory and a 32-GB solid state hard drive. This particular system is ideally suited for robotics purposes because of its light weight, small form factor, and low power consumption (less than 50 watts). Progress with LabVIEW was rapid, and within several weeks we created a rough program that replicated the performance of the old system.

We have not entirely abandoned the use of microcontrollers on the rover. We are currently experimenting with a LabJack UE9. It is a network-interfaced IO device that we plan to use alongside the computer to access peripherals. This will provide the ability to offload basic control tasks from the computer, allowing it to focus on work such as video processing without interfering with vital control functions. As an added benefit, it is readily controlled using LabVIEW, allowing us to integrate the two devices into one control program.

2. Motor Protection

It quickly became apparent that we would need to implement some levels of protection into the motor control system to protect the rover from rapid changes in drive commands. The previous rover utilized planetary gearboxes, allowing the speed controller to utilize motor braking to slow the rover whenever a stop command was sent. Despite operator training, we experienced gearbox damage as a result of operator mistakes, particularly sudden reversals. We desire to prevent such occurrences.

These changes became even more necessary as a result of the new rover drive train, which includes a custom gearbox using worm gears. In this system, the wheel must stop as soon as the motor does, exerting substantial torque on the gears. Since motor braking was not an option, we decided that we needed a function to smooth out, or soften, any rapid changes in motor speed.

The next decision was whether this algorithm should be implemented in the base station code or on board the rover. In the past, the base station took joystick input and determined what the motors should do. We realized that if there was ever an interruption in communications, the intermediate instructions to soften movement could potentially be lost, resulting in exactly the type of situation we wished to avoid. As a result, we determined that this code must run on the rover itself.

3. Motor Fail Safes

A related problem is the tendency for the rover to continue moving in a direction until it is told otherwise. If the rover ever loses communication, we do not want it to continue driving further out of range. The solution to this problem must also be implemented rover-side.

For hardware, we decided to replace the Mamba Max speed controllers with VEX Victor Pro 884 controllers. They have proved to be much easier to work with than the Mamba Max controllers. We have

decided to continue using the SSC-32 controller because of our past success with it and the ease of interfacing with it through LabVIEW.

D. Detail Design

1. Programming Process

In all of our detail designing and programming, we followed an orderly process of writing small modules, called Virtual Instruments (VIs), which perform specific tasks in an attempt to make the code easy to understand and modular. The following steps were taken in putting together the code:

Control-Side

- Read and display joystick data
- Concatenate joystick data into a comma-separated string and then parse it back into the original data to simulate sending those values over a network
- Add User Datagram Protocol (UDP) support to send values over network

Rover-Side

- Move a single servo based on a dial
- Move 4 servos
- Move standard motor based on joystick
- Calculate left and right wheel speeds based on commands
- Display commands received via UDP
- Execute commands
- Implement command softening

A test program with input controls and displays was written for each VI to confirm that it performed as expected. We caught many errors this way, and it simplified the process of combining the VIs into a single program. As we followed this process, we were able to make significant improvements in the rover-side computing, joystick converting, and control softening.

2. Rover-Side

The rover-side program consists essentially of two loops: a network communication loop and a motor control loop. The network loop is responsible for receiving instructions from the base station via UDP while the motor control loop sends them to the motors. We initially

tried combining both functions into one loop, but we discovered that delays were introduced in executing the commands that resulted in the network loop falling behind. Separating the two functions allows the rover to have the most recent command available for execution continually while allowing the motors to execute commands at their own rate without slowing down the rest of the program. We present an outline of this code below:

Network Loop

- Listen on UDP port
- Store command as a string in a single element queue
- Repeat every 100 ms

Motor Control Loop

- Read in string from queue
- If string is empty (no command received), generate string representing 'Stop'
- Parse string into X, Y, and Z joystick axis data
- Convert joystick data into Left/Right motor commands
- Soften reaction by limiting maximum magnitude of acceleration
- Send commands to motors
- Repeat when ready

3. Joystick Conversion

We considered a variety of schemes for converting joystick commands into motion. We eventually settled on using the same set-up utilized in the old rover because of its simplicity and natural behavior: Forward and backwards on the joystick act as expected, while moving it to an angle makes the rover veer slightly to that side. Additionally, twisting the joystick makes the rover spin in place.

Pseudo code

- If X or Y are above a threshold limit: Use standard steering formula:
 - $\text{Left} = -y \cdot (x + 1) / (1 + |x|)$
 - $\text{Right} = y \cdot (x - 1) / (1 + |x|)$
- If the joystick is centered (X and Y close to 0) and the joystick is twisted, spin in the direction of the twisting by instructing the

left wheels to move in the opposite direction of the right wheels.

Note: This code is designed for a four-wheeled rover. Since this year's model will have six wheels, two of which are further from the center of the rover than the other four, adjustments will have to be made to spin those wheels at slightly different velocities to compensate.

4. Softening

The purpose of this module is to ensure that the motor never undergoes excessive acceleration. Central to its operation is the concept of feedback: Each command set is influenced by the previous. An outline of this code follows:

Pseudo code

New command: command actually sent to motors

Desired command: input from operator

Previous command: previous command sent to motors

- Read in desired speeds
- Check if the difference between the previous command and the desired command is above a threshold (user adjustable)
 - False: New Command = Desired Command
 - True: New Command = $\text{Sign}(\text{Desired Command} - \text{Previous Command}) * (\text{Threshold}) + \text{Previous Command}$
- Previous Command = New Command

This system works very well. The threshold can be varied to allow for greater or less responsiveness. Testing will determine how high we can safely go with our current set-up. Currently, it is set at a conservative value of 2.

The softening also helps implement the fail-safe stop gradually: When no command is received from the network, a command of (0, 0, 0) is sent to the softener, slowing the rover down. If the lack of a command is due to a very short interruption of communications, the actual rover speed will only have been slightly reduced before a new command is received with the proper speed. Thus, momentary interruptions are filtered out while longer lasting interruptions cause the rover to stop.

5. Built-in VIs used:

We made use of the built in joystick interface, UDP communication, and SSC-32 servo controller abilities of LabVIEW.

E. Summary

We now have a functional control system for the rover. Using LabVIEW has greatly accelerated our rate of progress. We look forward to adapting the code as we refine the rover and add the code necessary to control peripherals such as a robotic arm and camera gimbal as they are completed.

V. Video and Communications

A. System Requirements

The rover must have a reliable communication link in harsh operating environments. This would include quality video and sensory feedback at a minimum range of 1 km and effective operation in non-line-of-sight (NLOS) situations. All communications and video equipment must be resilient to shock and high temperatures.

B. Previous Work and Observations

1. Problems with Low-Power Transmitters and Multiplexed Signals

One system utilized in the past was a multiple camera set-up relayed back to the base station using 200-mW analog transmitters on 910 MHz. To simultaneously transmit four signals, Mini-Circuits ZB4Cs-960-12W 4-1 multiplexers were employed to combine the signals onto one antenna. Problems were encountered with this system in NLOS situations and over longer distances. These problems were results of the low transmit power of the transmitters and the use of multiplexers, which degraded the signal by up to 6.0 dB.

2. Digital Video Limitations

We have also experimented with a purely digital system consisting of an onboard computer on the rover to compress the live video feed before transmission. The computer consisted of an Advantech PCM-9582F motherboard with a Celeron M 1.4GHz processor and 512-Mb RAM running Linux. Video was compressed using a Euresys Picolo Diligent video capture PCI card in DivX codec DX50. This system allowed for the use of multiple cameras over a single communica-

tions radio, which was also used to send control and sensory information. Unfortunately, low bandwidth forced us to limit the video bit rate to avoid delaying other vital control systems. A base station computer, consisting of a 2.0-GHz processor with 1 GB of DDR ram running Linux, was also insufficient in decoding the incoming video signal correctly and would drop frames to catch up with the feed. We also had many problems getting Linux to interface correctly with the video card on the rover, causing random crashes of the software. As a result, the video had to be limited to 352×240 pixels to get a full 30 frames per second.

3. Custom Set-up of Radios Necessary

The radios used on the rover have been a set of RouterBoard RB133s with Ubiquiti XR9 radio cards. These radios were paired with a set of standard wireless routers and set up as a transparent bridge. In this set-up, the routers handled all network traffic and their private networks were merged through the radio bridge.

C. Concept Generation

Reliable wireless performance is a combination of transmission power, antenna design, range, and noise. Communications on the rover can be split into two categories: essential and nonessential. Essential items, by definition, are systems whose failure would stop the rover from returning to the start gate by its own power. This primarily includes the drive commands and any sensory feedback necessary to control the rover. Nonessential items, while important, do not jeopardize this basic operation of the rover. These include, among other things, control of additional cameras, sensors, and task-specific instrumentation. The loss of functionality of nonessential items could prevent a task from being completed but will not incapacitate the rover.

The breakdown into these two categories was the primary means by which a communications system was developed. The most essential items necessitate the most reliable systems for their communication, while nonessential items do not need to be as robust.

D. Detail Design

To meet our requirements, several factors need to be considered. First, to simplify operation of the rover, we will use only unlicensed radio frequencies as outlined by the FCC, which include 902-928 MHz, 1.2GHz, 2.4-2.5 GHz, and 5.8-5.9 GHz. Free-space path loss, calculated using Eq. 1 (below), for our desired transmission distance of 1 km

yielded the following: for 900 Mhz, 91.5 dB loss; for 1.2 GHz, 94.0 dB loss; for 2.4 GHz, 100.0 dB loss; and for 5.8 GHz, 107.7 dB loss. These calculations are made assuming no Fresnel effect, no antenna gain, no loss associated with hardware imperfections, and good line of sight. While these calculations are idealized, it nonetheless gives us a reasonable approximation of the actual loss over the given distance. Based on these results we chose 900 MHz as our primary frequency, while 2.4 GHz was chosen as a secondary frequency over 1.2 GHz because available technologies are currently more advanced and accessible in the 2.4-GHz range.

Several options were considered in radio selection. It was decided that, as a result of past problems from combining video and control signals on a single radio, separate radios would provide us better performance and a more reliable set-up. Given the event of a radio failure, this would also provide us with a redundancy. It was also decided that frequency-hopping spread spectrum (FHSS) radios would be used for our 900-MHz controls/sensory feedback radio to maximize performance and reliability. FHSS radios offer us several advantages over a fixed-frequency radio. They rapidly hop between channels, providing for a higher resistance to background noise, the ability to operate on the same frequency as other radios with minimal effects, and high resistance to narrowband interference.

We have chosen to use a hybrid digital-analog system to meet our video needs. The system will consist of a digital set-up, allowing us to run multiple video feeds, with one camera on an analog transmitter as a back-up.

The base station communications set-up will consist of a Linksys WRT54GL router to interface multiple computers to our system and manage network traffic. Our primary radio is a FreeWave HTP-900re, which has an effective data rate of 570 Kbps. This FHSS radio will handle all critical functions including controls, GPS, and sensory feedback. This will be paired to an HG915G-NF15dB air grid dish antenna, which has a 30° horizontal beam width. The antenna will be connected to the radio with 1.5 ft of LMR-400 cable and mounted on a 35-ft antenna mast. The rover will have an equivalent radio set-up paired with a 10-dB omnidirectional antenna mounted at 4 ft. This set-up gives us the following results for our critical systems radio:

Transmit frequency: 900 MHz
Distance: 1 km
Radio TX power: 870 mW
TX Cable loss: 16 dBm
TX Antenna: 15 dB
RX Sensitivity: -96 dBm

RX cable loss: 6 dB

RX antenna: 10 dB

For further analysis, we use the following formulas:

$$\text{Free space loss} = 32.45 + 20\text{Log}_{10}(\text{freq}) + 20\text{Log}_{10}(\text{dist in km}) \quad (1)$$

$$\text{RX power} = \text{margin} - \text{RX sensitivity} \quad (2)$$

$$\text{Theoretical margin} = \text{TX power budget} + \text{RX power budget} - \text{free space loss} \quad (3)$$

$$\text{System operating margin (SOM)} = (\text{theoretical margin}/\text{TX power budget}) * 100 \quad (4)$$

Equations (1)-(4) yield:

Free space loss = 91.5 dB

Total TX = 28.4 dB

Total RX = 100 dB

Total RX Power = -59.1 dBm

SOM = 129.9%

Theoretical Margin = 36.9 dB

These results suggest that our system will have sufficient power to overcome our distance requirement and possible NLOS situations. Preliminary testing of the set-up provided very satisfactory results. With the base station mast at 10 ft, we obtained 94% successful transmission at an angle of 0-40° from the antenna's center in a NLOS scenario at more than 200 m. We are currently undergoing further testing, but these results leave us confident that the communication radios system will suffice for our needs.

E. Summary

From our research, we have developed a system for critical communications based on FHSS technology that will perform in the worst possible competition circumstances. We continue to work on establishing a comparable system to be used for our video feedback to ensure a stable communications system in spite of harsh operating conditions.

VI. Conclusion

In summary, there are many factors to consider in the development of robotic assistants to be used in extreme, non-Earth environments. These rovers will be instrumental in assisting our future

extraterrestrial bases in exploration and performing tasks to aid in their upkeep. Through our research and test experience, we are developing the technologies for such rovers. By applying what we have learned, we have designed a practical rover ready to compete in this year's University Rover Challenge, the design and technology of which has progressed our knowledge of feasible design concepts to be implemented in future Mars rovers.

Acknowledgments

We would like to thank the entire BYU Mars Rover Team, specifically, Mark Davis for extensive research on suspension systems, Morgan Gillespie for much of the CAD, and Ezekiel Merriam for leading the development of the arm. We would also like to thank our faculty advisor, Dr. David Allred for his continued support of our project and Dr. Doran Wilde for his technical advice and guidance in selecting a controls system. We would like to acknowledge the Brigham Young University Physics, Mechanical Engineering, and Electrical Engineering Departments for the funding and advisement to continue this research.

References

¹Sloan, Kevin, "Requirements and Guidelines for the 2011 University Rover Challenge." *Mars Society*. URL: urc.marssociety.org [accessed 7 March 2011].

²Harrington, B.D. and Voorhees, C., "The challenges of designing the rocker-bogie suspension for the Mars Exploration Rover," *Jet Propulsion Laboratory, NASA*, May, 2004, pp 1-3.

³Greco, C.R., "Real-time forward urban environment perception for an autonomous ground vehicle using computer vision and lidar," Masters Dissertation, Brigham Young University, 2008.

Document Flash Thermography

Cory A. Larsen and Gene A. Ware
Utah State University

Abstract

This paper presents an extension of flash thermography techniques to the analysis of documents. Motivation for this research is to develop the ability to reveal covered writings in archaeological artifacts such as the Codex Selden or Egyptian Cartonnage. An emphasis is placed on evaluating several common existing signal processing techniques for their effectiveness in enhancing subsurface writings found within a set of test documents. These processing techniques include false colorization, contrast stretching, histogram equalization, median filtering, Gaussian low-pass filtering, layered signal reconstruction, and thermal signal reconstruction, several contrast image definitions, differential absolute contrast, correlated contrast, derivative images, principal component thermography, dynamic thermal tomography, pulse phase thermography, fitting-correlation analysis, Hough transform thermography, and transmission line matrix fitting algorithm. The ability of flash thermography and the combined techniques to reveal subsurface writings; document strikeouts; and enhance surface inks will be evaluated. In addition, the differences in flash thermography parameters are evaluated for most effective imaging of the different document subsets.

I. Introduction

Flash thermography, a subset of pulsed thermography or pulsed video thermography, is a technique commonly used for nondestructive testing and evaluation in a variety of materials, including concrete [1] [2] [3], high-density polyethylene [4], aerospace composites [5] [6], wood and wood-based materials [7], and adhesive bond evaluation [8] [9]. This technology has yet to be applied to documents or archaeological artifacts to reveal covered writings such as those found in the Codex Selden or Egyptian cartonnage [10]. This paper develops the theory and application of flash thermography to documents and lays a foundation for the application of this technology to archaeological artifacts in general. Development of flash thermography for this application will provide the capability to nondestructively reveal covered writings—advancing the knowledge about ancient cultures without damaging irreplaceable artifacts.

Motivation for this research derives from the desire to analyze ancient archaeological documents with a nondestructive approach. Specifically of interest is the imaging of subsurface writings that may be obscured with a layer of some type of material. Ancient documents this technology could be applied to include, but are not limited to, palimpsests from the Roman Empire, Mesoamerican codices, and Egyptian cartonnage. Also of interest is use of flash thermography to possibly detect textual changes, such as strikeouts, where older ink writings are covered with a more recent layer of ink writing.

II. Prior Research

Several imaging techniques have been developed to enhance surface writing and underwriting contained in archaeological artifacts. Infrared (IR) reflectography has been used in the analysis of paintings [11] [12] and papyrus (cartonnage) [13]. The main application of IR reflectography is to see underdrawings beneath a layer of paint. With IR reflectography, a constant light source is used to excite the material. An IR imager captures an IR reflectogram detailing the different optical properties of the overlaying paint and the underwriting [14]. In addition to IR reflectography, transient thermographic techniques, including pulsed thermography, have also been used to analyze paint layers in artwork such as frescoes [15] [16] and general artwork [17] [18] [19] [20]. A comparative study was performed comparing pulse thermography, lateral heating thermography, and modulated thermography for the analysis of frescoes [16]. Pulsed thermography was shown to be successful in areas where X-radiography, IR reflectography, and ultraviolet (UV) examination had been unsuccessful [17]. Pulsed thermography

has not yet been applied to the evaluation of ancient documents or other types of archaeological artifacts.

A common technique for analyzing ancient documents is multi-spectral imaging (MSI). MSI has been shown to be effective for enhancing contrast between underwriting, overwriting, and the document substrate [21] [22]. MSI is performed by imaging documents in narrow spectral bands of light, allowing the spectral signature of the different document materials to be evaluated. Processing techniques for MSI images include, but are not limited to, the use of Markov random fields [21], spectral clustering [23], principal and independent component analysis [22], and linear spectral mixture analysis [22] [24].

The use of MSI has been successful in revealing obscured writing on the Archimedes palimpsest [24], carbonized scrolls [23], oxyrhynchus papyri [25], and the Dead Sea Scrolls [26] [27]. MSI is most effective enhancing writing that appears on or near the surface of the document. For example, the effectiveness of MSI to reveal the under codex within the Codex Selden was shown to be limited [28]. Another technique currently under investigation is X-ray fluorescence (XRF) imaging [29].

III. Description

Active thermography is an effective tool in nondestructive evaluation. Active thermography includes modulated (lock-in), pulsed, stepped, or vibro-thermography [30]. This work limits its investigation of flash thermography, a subset of pulsed thermography, as applied to the analysis of documents.

Flash thermography works by thermally exciting a surface with a flash-lamp discharge followed by using a high-speed IR camera to create a stream of the transient surface temperature images [31]. The surface temperatures cool as the heat is transferred through the document; however, when a material defect is reached, a portion of the heat is reflected back to the surface. This creates a relatively “warm” (or “cold”) spot on the surface that can be detected with the IR imager. Deeper subsurface detection is achievable over IR reflectography or MSI and can be applied to more general documents. Signal processing techniques are used on the video stream of flash thermography images to enhance defect contrast and determine quantitative parameters within the document.

The application of flash thermography to the analysis of documents has not been reported in the literature. This paper reports the application of flash thermography to documents and an analysis of the

effectiveness of current processing techniques when applied to documents.

IV. Experiment Methodology

A. Test Document Construction

The test documents were constructed to simulate ancient documents and were constructed out of a combination of materials. Substrate layers of card-stock and papyrus were used. Three ink types were evaluated: carbon based, iron gall, and ball-point pen. Finally, the ink layers were covered with either a substrate layer, a paint layer, or a mineral gesso mixture. For evaluating strikeouts, the base layer of ink was then covered with an additional layer of ink of the same type.

B. Equipment Used

The high-speed, mid-IR camera used was a Lockheed Martin/Santa Barbara Focal plane model SBF 180 with a custom data collection computer and software. The flash units were SunPak Pro-System 622 Super. The camera was set horizontally and aimed at an angled mirror. This mirror was used to image the test document lying flat on the surface of a table. The flash units were then elevated above the sides of the document on the table with the flash heads pointed toward the test document. Two camera lens notch filters were used, a 3.42–4.05 μm and a 2.65–3.24 μm filter were evaluated. A sampling frequency of 87 Hz was used for the acquisition of all images.

V. Processing Techniques

A. Background Theory

The majority of the algorithms herein presented process the image data on a pixel by pixel basis, evaluating the time series of each pixel separately without accounting for lateral diffusion. These time series represent the post-flash surface temperature decay of the material through time. It is often assumed that the diffusion into the material is significantly greater than the lateral diffusion and therefore the lateral diffusion can be neglected. This allows the diffusion into the document to be described using the equation for one-dimensional thermal diffusion as given by

$$\frac{\partial^2 T}{\partial x^2} = \frac{1}{\alpha} \frac{\partial T}{\partial t}, \quad (1)$$

where T is the temperature and α is the thermal diffusivity of the material. For an ideal impulsive heat flux, the response for a semi-infinite surface is [32]

$$T(x, t) = \frac{Q}{e\sqrt{\pi t}} e^{-\frac{x^2}{4\alpha t}}, \quad (2)$$

where $e = \sqrt{k\rho c}$ is the thermal effusivity of the material as determined by the thermal conductivity, k , mass density ρ , and specific heat c . Q is the quantity of energy absorbed by the surface. Time is represented by t and the depth into the material is given by x . Since the thermal imager can only respond to surface temperatures, Equation (2) is evaluated at $x = 0$, resulting in the surface temperature decay modeled by

$$T_{surf}(t) = T(0, t) = \frac{Q}{e\sqrt{\pi t}}. \quad (3)$$

However, the thermal imager only gives relative temperatures, therefore let $T_{surf}(t) = T_{surf}(t) - T_{ambient}$, where $T_{ambient}$ is the pre-flash initial temperature of the sample.

The response is thus more accurately described relative to the thermal imager as

$$\Delta T_{surf}(t) = \frac{Q}{e\sqrt{\pi t}}. \quad (4)$$

Equation (4) provides a basis for many of the algorithms discussed in this section. This equation can be further simplified to [31]

$$\Delta T_{surf}(t) = T_{init} \sqrt{\frac{T_s}{t}}, \quad (5)$$

where T_{init} is the value of the surface temperature at one time step, T_s , and is given by

$$T_{init} = \frac{Q}{e\sqrt{\pi T_s}}. \quad (6)$$

The techniques described throughout this section were implemented in MATLAB 2010b.

B. Pseudo-Color Images

Human eyes contain two classes of light receptors: cones and rods. Cones are highly sensitive to color and are able to discern smaller changes in intensity than rods. This is due to the fact that cones, unlike rods, are connected to an individual nerve ending [33]. Therefore, it is useful to provide pseudo-color to the gray-scale intensity images obtained from the thermal cameras. These pseudo-color images are obtained through a process of intensity slicing and color coding. Intensity slicing is a technique used to separate different intensity values in a gray-scale image. Given a gray-scale image with L intensity levels, the image can be separated into P intensity levels where $0 < P < L - 1$. By color coding the intensity values between regions, a pseudo-color image is created [33]. To avoid creating a mosaic effect within the image, it is better to use a continuous color map; however, in some situations, choosing a discrete color mapping scheme may enhance the visibility of a particular feature [30]. Pseudo-color images can be created in MATLAB using the `colormap` command using the *jet* color map parameter.

Pseudo-colorization can be used as an aid to help enhance visualization of defects. Pseudo-colorization is often used as a pre-processing technique to analyze the data before processing. If no defects can be detected, even subtly, from a visual inspection, then it is usually an indicator that the processing algorithms will not be able to effectively enhance any under writing. Pseudo-color images can also be used to aid in visualization of the post-processed images.

C. Contrast Stretching

Contrast stretching is a point processing technique used to expand the dynamic range of an image to increase visibility of image features. This technique is used to enhance visibility of the raw thermal images or as a post-processing step for the other techniques described in this section. The simplest form of contrast stretching is image normalization, which is given by

$$I_N = (I - c) \left(\frac{b - a}{d - c} \right) + a, \quad (7)$$

where I is the input image with initial range $[c, d]$ and I_N is the normalized output image in the desired range $[a, b]$. Note that the normalization process is used on individual frames within the time sequence. Image normalization is greatly affected by dead pixels and other out-

liers in pixel values. To compensate, a useful technique is to saturate the top 1% and bottom 1% of pixels. This can be implemented with the MATLAB function `imadjust`.

D. Histogram Equalization

A histogram of a digital image is a discrete function that counts the number of pixels within given intensity level bins. The histogram is a discrete estimation of the probability density function (PDF) of the image. In histogram equalization, the goal is to transform the histogram of the input image into an output image with a uniformly distributed histogram. This technique enhances the global contrast of an image by adjusting the distribution of intensities within an image. If a defect lies in a region with low local contrast, histogram equalization will increase the contrast of that region allowing for enhanced defect visualization.

The transformation on each pixel intensity value is represented by

$$s = T(r), \quad (8)$$

where s is the equalized pixel value, r is the input pixel value, and $T(r)$ represents the transformation performed on r to obtain s . The transformation is given by [33]

$$s = T(r) = (L - 1) \int_0^r p_r(w) dw, \quad (9)$$

where L represents the number of intensity levels and w is a dummy variable of integration. Equation (9) represents the cumulative distribution function of the image. It is proved elsewhere that this transform results in a uniform distribution [33]. For discrete digital images, the transformation function becomes

$$s_k = T(r_k) = \frac{L - 1}{MN} \sum_{j=0}^k n_j \quad k = 0, 1, 2, \dots, L - 1 \quad (10)$$

where MN is the total number of pixels in the image, and n_k is the number of pixels that have intensity r_k . A plot of n_k versus r_k results in the histogram of the image [33]. Since the histogram is a discrete approximation of the PDF and no new intensity levels can be created, perfectly flat histograms are rare in practical images [33]. Histogram equalization can be implemented in MATLAB using the `histeq` function. The histogram equalization process is used to post-process the images after

running the other algorithms described in this section or to better evaluate the raw, unprocessed images.

E. Image Filters

Two image filters have been found to be useful in noise reduction in flash thermography data, namely a median filter [33] and a Gaussian low-pass filter [30]. Median filters are useful for removing salt and pepper noise. The median filter is a nonlinear filter that ranks the pixels within a neighborhood, replacing the center pixel with the median of the intensity values. The median filter has advantages over a mean filter because it is not affected by outliers (such as those caused by dead pixels) and better preserves edges within the image. The median filter is common in imaging software and can be implemented in MATLAB using the `medfilt2` function. The Gaussian filter is a frequency-domain filter that assumes the image has a limited bandwidth, and any spatial frequencies above the given bandwidth are the result of noise. Since image noise tends to be characterized by high spatial frequencies, a low-pass filter can be used to reduce the noise content in the image. A derivation of the Gaussian filter for thermal images can be found elsewhere [30]. The Gaussian filter can be implemented in MATLAB using the `fspecial` command to create the filter and the `imfilter` command to apply the filter to the image.

F. Synthetic Signal Reconstruction Techniques

Two synthetic signal techniques were employed to reconstruct the signal, namely thermal signal reconstruction (TSR) [34] [35] and layered reconstruction (LR) [36]. Advantages of using reconstructed, synthetic data include (1) significant improvements in sensitivity, (2) a reduction of blurring, (3) increased depth range, (4) decreased memory requirements, and (5) improvements in signal-to-noise performance. The TSR reconstruction processes each individual pixel's time sequence, rather than each frame as a whole. The time response data can be linearized by transforming to a logarithmic domain. The logarithmic transform of Equation (4) is

$$\ln(\Delta T_{surf}(t)) = \ln\left(\frac{Q}{e}\right) - \frac{1}{2} \ln(\pi t). \quad (11)$$

This implies that, regardless of the thermal properties of the material, the logarithmic decay response will be a straight line with a slope of $-1/2$ for an ideal, defect-free region. The linearized sequence can be modeled using a least-squares fit to a N th order polynomial

$$\ln[\Delta T_{surf}(t)] = \sum_{n=0}^N a_n [\ln(t)]^n. \quad (12)$$

It was found that a fifth or sixth order polynomial effectively acts as a low-pass filter to smooth the data without reconstructing the noise [35]. The data can be reconstructed using [34] [35]

$$\Delta T_{surf} = \exp \left(\sum_{n=0}^N a_n [\ln(t)]^n \right). \quad (13)$$

In the layered reconstruction (LR) [36] algorithm, a multilayered approach is taken. It is proposed that the signal can be reconstructed using the following equation:

$$T_{surf}(t) = T_f + \sum_{i=1}^j A_i e^{-\frac{t}{\tau_i}}, \quad (14)$$

where A_i is the amplitude of each exponential function and T_f is the steady-state temperature. The time constant of each layer is defined as $\tau_i = \frac{1}{\alpha_i B_i^2}$, where α_i is the thermal diffusivity of the layer and B_i is determined by the boundary conditions. This approach can be further simplified by introducing normalized variables. Let ζ be the normalized time defined as

$$\zeta = \frac{t}{t_f}, \quad (15)$$

where t_f is defined as the time it would take the temperature to diminish 99.3%, or 5 time constants. Let θ be the normalized temperature difference, defined as

$$\theta = \frac{T_{surf}(0) - T_{surf}(t)}{T_{surf}(0) - T_{surf}(t_f)}. \quad (16)$$

Equation (14) can be normalized using the definitions of ζ and θ as follows:

$$\theta(\zeta) = 1 - \sum_{i=1}^j \beta_i e^{-\frac{\zeta}{\psi_i}}, \quad (17)$$

where $\beta_i = \frac{A_i}{T_{surf}(0) - T_{surf}(f)}$ and $\psi_i = \frac{\tau_i}{t_{fi}}$. A least-squares method can be used to solve for the β_i and ψ_i of each layer, both of which are normalized between 0 and 1. The synthetic signal can be reconstructed using the fitted coefficients and Equation (17). In addition to performing a pure reconstruction, the synthetic signal can be modified. For example, the time points used to reconstruct the signal can be altered to effectively up-sample, down-sample, or interpolate the data [37].

G. Contrast Definitions

Four commonly used definitions of contrast are absolute, running, normalized, and standard contrast. Each of the techniques are outlined below and are summarized [30]. The absolute contrast is defined as the excess temperature over a defect-free region at a given time t and is given by

$$C_{abs}(t) = T_{defect}(t) - T_{defectFree}(t), \quad (18)$$

where T is the temperature over a defect region and a defect-free region, respectively. This increases the contrast and improves the visibility of the defective region over the defect-free region. The running contrast reduces the effects of differences in surface emissivities and is defined as

$$C_{run}(t) = \frac{C_{abs}(t)}{T_{defectFree}(t)}. \quad (19)$$

Note that if the contrast images are post-processed with the contrast stretching techniques given previously, then the absolute contrast and running contrast are the same. The normalized contrast can be computed with respect to the end of the thermal process, at time t_{end} , or the time of temperature max, t_{max} (for pulsed thermography, this is the first frame). The normalized contrast is defined as

$$C_{norm}(t) = \frac{T_{def}(t)}{T_{def}(t_{max})} - \frac{T_{defectFree}(t)}{T_{defectFree}(t_{max})}, \quad (20)$$

where t_{max} can be replaced with t_{end} .

Finally, the standard contrast was developed to eliminate contributions from the surrounding environment by subtracting pre-flash information given at time t_0

$$C_{std}(t) = \frac{T_{defect}(t) - T_{defect}(t_0)}{T_{defectFree}(t) - T_{defectFree}(t_0)} \quad (21)$$

Each of these contrast images requires selection of a defect-free area.

H. Differential Absolute Contrast (DAC)

The contrast methods described previously are greatly affected by nonuniform heating of the surface and requires selection of a defect-free area. Differential absolute contrast (DAC) [38] removes the need for manual selection of a defect-free area and is therefore more robust when nonuniform surface heating occurs. Let t' be the time at which the defect begins to appear in the sequence and $\Delta T(t)$ represent the frame at time t . Define the defect-free area as

$$\Delta T_{snd}(t') = \Delta T(t'). \quad (22)$$

From Equation (2), the value of $\frac{Q}{e}$ can be solved to obtain

$$\frac{Q}{e} = \sqrt{\pi t'} \Delta T(t'). \quad (23)$$

Using this result in Equation (2), the ideal defect-free area can be found as

$$\Delta T_{snd}(t) = \sqrt{\frac{t'}{t}} \Delta T(t'). \quad (24)$$

By the definition of the defect-free region combined with the definition of absolute contrast given in Equation (18), the DAC image is given by

$$DAC(t) = \Delta T(t) - \sqrt{\frac{t'}{t}} \Delta T(t'). \quad (25)$$

Since the input pulse is not an ideal impulse, small differences in pulse length can be accounted for within different parts of the image. Defin-

ing t_e to be the amount of error in pulse length, a fit can be made in the logarithmic domain to find t_e and compensate for the error. The error-compensated DAC image is then found as

$$DAC(t - t_e) = \Delta T(t - t_e) - \sqrt{\frac{t' - t_e}{t - t_e}} \Delta T(t' - t_e), \quad (26)$$

where the values of t_e can vary over every pixel. The definition of the defect-free area can also be used with the other definitions of contrast given in Section V-G. A technique called interpolated differential absolute contrast (IDAC) has been developed to remove the need for manual selection of time t' [39]. In addition, thermal quadrapole theory has been used to extend the validity of DAC to later times [40] [41].

I. Derivative Images

A defect can theoretically be detected without the use of a reference region by evaluating deviations from the ideal $-1/2$ slope seen in Equation (11). To increase defect contrast and limit the effects of blurring caused by lateral diffusion, the derivatives of the analytical model for the data are taken. The derivatives facilitate detection of an earlier time of maximum contrast, thus reducing lateral diffusion blurring; however, the diameter of the subsurface defect must be greater than its depth beneath the surface for the lateral diffusion to be effectively ignored [42]. This algorithm begins by reconstructing the signal using the TSR technique described in Section V-F to obtain an analytical expression for the data. The pixel time histories are differentiated using the expressions [35]

$$\frac{d \ln(\Delta T_{surf}(t))}{d \ln(t)} = \sum_{n=0}^N n a_n \ln(t)^{n-1}, \quad (27)$$

$$\frac{d^2 \ln(\Delta T_{surf}(t))}{d \ln(t)^2} = \sum_{n=0}^N n(n-1) a_n \ln(t)^{n-2}. \quad (28)$$

The reconstructed signal and its time derivatives are transformed into the linear time domain through Equation (13) and by taking the exponent of the function as shown

$$\frac{d \Delta T_{surf}(t)}{dt} = \exp \left(\sum_{n=0}^N n a_n \ln(t)^{n-1} \right), \quad (29)$$

$$\frac{d^2 \Delta T_{surf}(t)}{dt^2} = \exp \left(\sum_{n=0}^N n(n-1) a_n \ln(t)^{n-2} \right) \quad (30)$$

The resulting derivative time sequences can be outputted for analysis. Quantitative defect depth analysis can be estimated as described by Omar et al. [43] and others [31] [44].

J. Principal Component Thermography (PCT)

Principal component thermography (PCT) [45] [46] uses singular value decomposition (SVD) to reduce data to a compact statistical representation of the spatial and temporal variations relating the contrasts associated with underlying material defects [45]. In flash thermography data, a series of two-dimensional image frames are stored sequentially, creating a three-dimensional data set. To perform PCT, a raster-like operation must be performed to create a two-dimensional representation of the three-dimensional data. Given that the original data are loaded into an image cube with dimensions $N_x; N_y;$ and N_t , where the N_x and N_y describe the pixel dimensions of each frame and N_t describes the number of frames. This image cube is then transformed into a matrix A with dimensions $M \times N_t$, where $M = N_x N_y$. The column vectors of M are standardized to correct for individual detector pixel characteristics. This standardization is achieved through

$$\hat{A}(n, m) = \frac{A(n, m) - \mu_n}{\sigma_n}, \quad (31)$$

where

$$\mu_n = \frac{1}{N_t} \sum_{m=1}^{N_t} A(n, m), \quad (32)$$

$$\sigma_m^2 = \frac{1}{N_t - 1} \sum_{n=1}^{N_t} (A(n, m) - \mu_n)^2. \quad (33)$$

Any $M \times N$ matrix can be decomposed through SVD into the following elements

$$A = U \Gamma V^T, \quad (34)$$

where Γ is a diagonal matrix containing the singular values of matrix A , and U and V^T contain the left and right singular vectors of A . In this

application, the matrix U contains a set of orthogonal basis functions that describe the spatial variations within the data and the matrix V^T contains the corresponding characteristic time behavior, which can be used to estimate defect depths. By reversing the raster transformation applied to create A on U , the empirical orthogonal functions (EOF) of the data are produced.

An analysis was performed showing that the first two modes tend to contain 99% of the variance within the data, although some leakage occurs in the higher-order modes [46]. The first mode describes a response similar to that of a uniform slab; however, the second mode, characterizes a nonuniform field created by material anomalies, and therefore has been named the primary contrast mode (PCM). A drawback of the original PCT algorithm is that it may enhance some defects at the cost of other defects [47]. The flaw depth may be estimated from the second principal component (PC), contained in the matrix V^T , and by knowing the thermal diffusivities of the material using the technique developed in [45] [48].

K. Dynamic Thermal Tomography (DTT)

Two versions of the DTT algorithm were implemented, those referred to as classical DTT and as reference-free DTT [47] [49] [50]. The classical algorithm requires the selection of a reference, defect-free region. The frames are first normalized by dividing each frame with the first postflash frame. The difference between the time sequence of each pixel and the reference signal is taken as

$$\Delta T(x, y, t) = T(x, y, t) - T_{ref}(t). \quad (35)$$

Two images can be created from the new ΔT signal: (1) an image of the maximum value of ΔT ; and (2) the time at which the maximum value occurs. These images are referred to as a maxigram and a timegram, respectively. This creates a synthetic image that is sampled at each pixel's "optimal" time. The corresponding transit times in the timegram can be used to create tomographic slices of the document.

The reference-free approach is similar to the classical approach, but removes the need for manual selection of the defect-free region. Instead, different-order polynomials are fitted to the normalized temperature response. The low-order polynomial will only reflect general behavior of the material, whereas higher orders will include the behavior of defects. In this case, a third- and a sixth-order polynomial were used. The difference for each pixel can be found through

$$\Delta T(x, y, t) = T_h(x, y, t) - T_l(x, y, t). \quad (36)$$

The maxigram and timegram are obtained from $\Delta T(x, y, t)$ and outputted. Defect depths can be found using procedures described elsewhere [51].

L. Pulse Phase Thermography (PPT)

PPT is a combination of two forms of thermography: flash (pulse) thermography and modulated thermography. In flash thermography, a pulse of heat energy is deployed into the target, and the transient decay of the resulting surface temperature is analyzed. Alternately, in modulated thermography, the target is subject to a sinusoidal temperature stimulation in which standing thermal waves are created within the material. These standing thermal waves are analyzed by their magnitude components and phase shifts with respect to the reference modulation. The magnitude images are proportional to local optical and infrared surface features; however, the phase shift images are relatively independent of these features. As a result, the phase images can probe roughly twice the thickness given by the magnitude image and are therefore the output of interest [52].

PPT uses the principle that a pulse of energy in the time domain contains all frequencies in the frequency domain. Since the input pulse is not an ideal delta function, but rather a rectangular pulse, the resulting frequencies are given by a sinc function [30]

$$F(f) = A_p T_s \text{sinc}(\pi f T_s), \quad (37)$$

where f is the frequency variable, A_p is the pulse amplitude, and T_s is the sampling rate. In effect, all frequencies are analyzed simultaneously in PPT rather than at a single frequency as in modulated thermography [53]. In PPT, the discrete Fourier transform of each pixel's time series is computed using

$$F(u) = \frac{1}{N} \sum_{n=0}^{N-1} h(x) e^{-j2\pi u x/N} = R(u) + jI(u), \quad (38)$$

where $R(u)$ and $I(u)$ are the real and imaginary parts, respectively, of the transformed sequence, $F(u)$. The magnitude and phase responses are obtained from the transformed data through

$$\phi(u) = \tan^{-1} \left(\frac{I(u)}{R(u)} \right), \quad (39)$$

$$|F(u)| = \sqrt{R(u)^2 + I(u)^2}, \quad (40)$$

resulting in a series of magnitude and phase difference output images.

As previously stated, the phase images are usually of interest because of their relative independence of medium surface features. The resulting series of images correspond to frequencies ranging from 0 to $1/\Delta t$, where Δt is the time interval between images. The lower the frequency, the deeper the image is able to probe. For the phase offset images, it was found most useful if the maximum phase offset, ϕ_{max} , was found for each pixel time history and outputted into a single resulting image [53]. Another form of PPT is computed using the Wavelet Transform (WT) in place of the Fourier Transform (FT). The advantage is that wavelets preserve time information of the signal and are correlated to defect depth, allowing quantitative evaluations [54] [55]. Another technique uses the Hough Transform to retrieve the blind frequencies, which are correlated with the defect depth [56]. It was also found that preprocessing the images with the reconstruction technique given in TSR improved the depth resolution of PPT [57].

M. Correlation Images

The fitting-correlation algorithm (FCA) [58] begins by reconstructing the signal using the technique specified in TSR (see Section V-F). The reconstructed signals are then evaluated to see how closely they match either an “ideal” signal or a manually selected defect-free signal. The ideal signal can be found using Equation (5) or Equation (24). Two methods are used to evaluate the closeness of the fit: the correlation coefficient and the angle cosine. The correlation coefficient is calculated as

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (41)$$

and the angle cosine as

$$\cos \theta = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}}. \quad (42)$$

The resulting correlation coefficient image and the angle cosine image are output.

The correlated contrast technique [59] builds upon the FCA algorithm. Klein argues that the chosen reference signal is irrelevant because differences can be adjusted for by changing the color map. In addition, it is noted that the dynamic range of the correlation coefficient image can be large, thus logarithmic color palettes are more appropriate than linear color palettes for visualizing the images [59]. The correlated contrast implemented in the IR View toolbox uses a fifth-order logarithmic root.

N. Transmission Line Matrix Fitting Algorithm (TLMFa)

The transmission line matrix is a numerical technique commonly used to model voltages transmitted within transmission lines and has been extended as a method to model forward diffusion. Transmission lines can be modeled using the lossy wave equation, or the telegrapher's equation, as given

$$\nabla^2 v = L_d C_d \frac{\partial^2 v}{\partial t^2} + R_d C_d \frac{\partial v}{\partial t}, \quad (43)$$

where $v = v(x; y; z; t)$ and represents the voltage, and R_d , C_d , and L_d are the distributed parameters of a lumped transmission line and are the resistance, capacitances, and inductance. Now to begin applying this to thermal diffusion, interpret R_d and C_d as the thermal resistance and capacitance per unit length, respectively. Also let the voltage be analogous to the temperature $v = u$. By making appropriate space and time discretizations, the inductance term will disappear, reducing Equation (43) to

$$\nabla^2 u = R_d C_d \frac{\partial u}{\partial t}. \quad (44)$$

By defining $\alpha = \frac{1}{R_d C_d}$, the equation above can be rewritten as

$$\nabla^2 u = \frac{1}{\alpha} \frac{\partial u}{\partial t}, \quad (45)$$

which, letting α represent the thermal diffusivity, is the thermal diffusion equation describing heat flow. Thus, given small enough space and time discretizations, thermal diffusion can be modeled using the telegrapher's equation where the voltage is equivalent to the temperature

and the current is equivalent to the heat input. The only errors introduced are those imposed by the space and time discretizations [60].

The TLM algorithm can be based off two models for a transmission line, the T-network or Π -network models. These lead to two formulations of the TLM algorithm, referred to as the link-line TLM node or link-resistor TLM node, respectively. Both methods achieve the same results but have different design approaches. The link-line nodal arrangement is derived in the following paragraphs. From the derivation above, the temperature is analogous to the voltage and the heat energy is analogous to the current. These analogies will be used for the rest of the derivation.

The transmission line can be seen as ideal segments of line with lumped resistance and capacitance. An impulse entering this line will

travel unimpeded for time $\frac{\Delta t}{2}$ at which point it encounters a discontinuity. Some of this initial pulse will transmit on and part will be reflected back. This is determined by the reflection and transmission coefficients, which are given by

$$\rho = \frac{R}{R + Z}, \quad (46)$$

$$\tau = \frac{Z}{R + Z}. \quad (47)$$

Assuming two input impulses are given approaching node x at time k .

These pulses can be represented as ${}^i_k V_L(x)$ and ${}^i_k V_R(x)$ (the i indicates that it is incident on node x). The voltage at the node is given by the sum of the left and right bound pulses

$${}_k \phi(x) = {}^i_k V_L + {}^i_k V_R. \quad (48)$$

The incident pulses are then scattered using the transmission and reflection coefficients given above

$${}_k^s V_L = \rho {}^i_k V_L + \tau {}^i_k V_R, \quad (49)$$

$${}_k^s V_R = \tau {}^i_k V_L + \rho {}^i_k V_R. \quad (50)$$

The scattered pulse then travel along the line until they become incident on adjacent nodes

$${}^{i}_{k+1} V_L(x) = {}^s_k V_R(x - 1), \quad (51)$$

$${}^i_{k+1}V_R(x) = {}^i_kV_L(x+1). \quad (52)$$

The processes of summation (48), scatter (50), and connect (52) are repeated to create the link-line TLM algorithm.

Using a least-squares fitting method, such as the Levenberg- Marquardt Method, the TLMFa algorithm can be applied to each individual pixels time response sequence and the reflection coefficients can be determined. Since the reflection coefficients are dependent on the thermal properties of the material, outputting images created from the each pixel's reflection coefficient for each node provides a characterization of a defect.

O. Hough Transform Thermography (PHTTa)

The Hough transform is a geometrical transform that is used in image processing to find geometrical structures within an image. The Hough transform is represented by

$$\rho = x \cos \theta + y \sin \theta, \quad (53)$$

where the points $(x; y)$ are transformed into the Hough space (ρ, θ) . The pulse thermography Hough transform algorithm (PHTTa) [61] [62] uses the Hough transform to evaluate thermal sequences removing the need for an operator by identifying pixels that follow the $-1/2$ slope shown in Equation (11). Each point is transformed pixel-wise into sinusoidal curves in the Hough space. The Hough space acts as an accumulator that sums the votes of all pixels in the sequence providing an indication of the points that follow a $-1/2$ slope. The points that correspond to a $-1/2$ lie in the $\theta \approx 1.1071$ rad column in Hough space [61]. By analyzing the distribution of values in the $\theta \approx 1.1071$ rad column, defective areas can be separated from nondefective areas. This is accomplished by performing a least-squares fit of an exponential function of the form

$$\zeta(x) = ae^{\left(\frac{x-b}{c}\right)}. \quad (54)$$

It has been shown that parameters a and c are highly correlated to the defect depth and images created from these parameters effectively enhance the contrast between the defect-free and defect regions [61].

VI. Summary and Conclusions

The conclusions developed from the experimental results are summarized in this section. The flash thermography parameters for

effective flash thermography data acquisition are outlined. A summary of methods for pre-processing, processing, and post-processing the data to increase ink visualization are given. The ability of flash thermography to image a variety of materials are discussed. Finally, this paper concludes with recommendations for further research and is used as a basis for the authors future investigations [63].

A. Thermography Data Acquisition Parameters

Several flash thermography parameters were evaluated during the acquisition of the data sets. Recommended values are discussed below.

Pulse Amplitude :

Four synchronized flash units were sufficient to reveal subsurface layers of ink (after processing) through paper, card stock, papyrus, and thin gesso layers. Large pulse amplitudes are desirable, but care must be taken to reduce the residual heating from the flash bulb that may occur. For document strikeouts or enhancing surface writing, a large pulse amplitude is unnecessary.

Pulse Duration

Residual heating from flash pulse tails were not found to have a strong negative effect on the detection process for deep-seated subsurface writings, provided the tail decayed relatively quickly and absorbed uniformly into the document. For document strikeouts and surface ink enhancement, a minimal input pulse length that closely approximates an impulse is desired.

Sampling Frequency :

A sampling frequency of 87 Hz was found adequate for subsurface defects; however, a faster sampling rate is recommended for document strikeouts and surface ink enhancement.

Lens Filters

The 3.42 - 4.05 μm filter greatly outperformed the 2.65 - 3.24 μm filter.

B. Pre-Processing

The following techniques were found useful for preprocessing flash thermography data when applied in the following order:

- 1) Subtraction of a preflash image to each frame to reduce effects of non-uniform surface emissivities;
- 2) A 3×3 pixel median filter applied to each frame for removing outliers caused by dead pixels;
- 3) A 3×3 Gaussian low-pass filter, with a variance parameter of 0.8, applied to reduce spatial image noise in each frame;
- 4) TSR, using a sixth-order polynomial fitting, for reducing temporal noise.

C. Processing

Effective processing techniques for each document class are presented. Table 1 outlines a rating system used in classifying the algorithms.

Table 1. Algorithm Rating System		
Value	Definition	Explanation
3	Highly recommended	Consistently gives great results with minimal computational expense.
2	Recommended	Performs well but inconsistently. May only work for simple document structures
1	Moderately recommended	Performs well occasionally or has a significant computation expense.
0	Not recommended	Either fails to perform well or computational expense is prohibitive.

Table 2 evaluates the pre-existing algorithm techniques used for processing flash thermography data of documents. Note that because of the similarities between processing strikeouts and surface ink enhancement, the ratings for strikeouts also apply to surface writing enhancement.

Table 2. Processing Algorithms Summarized				
Processing Technique (Section)	Sub-surface Rating	Strike-out Rating	Wall Time (Sec)	Comments
Contrast definitions (V-G)	1	1	0.0395 1	Can be used as a post-processor.
Differential absolute contrast (DAC) (V-H)	3	3	0.0319 1	Requires parameter selection.
Derivative images (V-I)	2	3	4.037	Use limited time sequence for strikeouts.
Principal component thermography (PCT) (V-J)	2	2	0.7850	May enhance some features over others. Use limited time sequence for strikeouts.
Classical dynamic thermal tomography (DTT) (V-K)	1	2	2.948	Performs moderately for detecting strikeouts.
Reference-free dynamic thermal tomography (DTT) (V-K)	2	1	6.843	Unable to reveal the deepest defects and has a moderate computational load.
Pulsed phase thermography (PPT) (V-L)	3	3	0.3465	Consistently performs well.
Correlation images (V-M)	0	0	10.76	Generally poor performance.
Transmission line matrix fitting (TLMFa) (V-N)	0	0	4044	Prohibitively large computational expense.
PT Hough Transform (PTHTa) (V-O)	1	0	125.9	Large computational expense.

D. Post-Processing

Essential post-processing methods include the following contrast stretching techniques:

- 1) Manual histogram manipulation,
- 2) Image normalization,
- 3) Image normalization with ends saturation,
- 4) Pseudo-color images.

Other techniques that may be useful, depending on the situation, include:

- 1) Histogram equalization,
- 2) Image filter (median or Gaussian),
- 3) Absolute and running contrast definitions.

E. Materials

This research demonstrated that flash thermography can be used to detect three types of inks: carbon based, iron gall, and modern ball-point. These inks were detected beneath several covering layers including card stock, papyrus, paint, a gesso mix, and the ink itself.

F. Significant Contributions

This research resulted in the following significant contributions.

- 1) A comprehensive analysis of existing pulsed thermography processing techniques applied to documents, the magnitude of which does not exist in the literature. Preexisting techniques investigated include pseudo-color images, contrast stretching, synthetic signal reconstruction, median and Gaussian low-pass filters, four contrast definitions, DAC, derivative images, PCT, classical and reference-free implementations of DTT, PPT, correlation images, TLMFa, and PTHTa.
- 2) A proof-of-concept for detecting subsurface ink writings, strike-outs, and surface ink enhancement in documents.
- 3) Provides an evaluation of flash thermography parameters for the most effective document imaging.

VII. Recommendations for Future Research

There are two principal directions suggested for future research. The first recommendation is the development of new algorithms specific to subsurface ink detection. New techniques could be developed through better modeling of the structure of the documents of interest. The second recommendation is research into alternate heating methods to facilitate for greater excitation energy. Methods of future research could explore step heating, thermal transfer heating, or modulated thermography techniques. For document strikeouts, it is recommended that the experiments be repeated using a thermal camera capable of significantly faster frame rates. Investigation of techniques that account for the lateral diffusion through the ink would also be beneficial.

VIII. Acknowledgments

Special thanks to Pedro Sevilla and James Peterson of the Space Dynamics Laboratory for the use of their equipment and helping with the acquisition of the thermal images. Thanks to Dr. Doran J. Baker of the Rocky Mountain NASA Spacegrant Consortium for funding the research and providing assistance as needed. Finally, thanks go to Dr. Todd Moon and Dr. Jake Gunther of Utah State University for sharing their advice and expertise.

References

- [1] F. C. Sham, N. Chen, and L. Long, "Surface crack detection by flash thermography on concrete surface," *Insight*, vol. 50, no. 5, pp. 240–243, April 2008.
- [2] J. Sham Fung Chu, "Studies of Using Infrared Flash Thermography (ft) for Detection of Surface Cracks, Subsurface Defects and Water-Paths in Building Concrete Structures," Ph.D. dissertation, City University of Hong Kong, July 2008.
- [3] H. Nayeb-Hashemi, D. Swet, and A. Vaziri, "New electrical potential method for measuring crack growth in nonconductive materials," *Measurements*, vol. 36, pp. 121–129, 2004.
- [4] M. A. Omar, Y. Zhou, R. Parvataneni, and E. Planting, "Calibrated pulse-thermography procedure for inspecting HDPE," *Research Letters in Materials Science*, vol. 2008, no. 186427, p. 4, 2008.

- [5] K. T. Wan and C. K. Y. Leung, "Fiber optics sensor for the monitoring of mixed mode cracks in structures," *Sensors and Actuators*, vol. 135, pp. 370–380, 2007.
- [6] C. Ibarra-Castanedo, M. Genest, P. Servais, X. P. V. Maldague, and A. Bendada, "Qualitative and quantitative assessment of aerospace structures by pulsed thermography," *Nondestructive Testing and Evaluation*, vol. 22, no. 2-3, pp. 199–215, June-September 2007.
- [7] P. Meinschmidt, "Thermographic detection of defects in wood and wood-based materials," 14th International Symposium of Nondestructive Testing of Wood, May 2005.
- [8] M. Y. Y. Hung, Y. S. Chen, S. P. Ng, S. M. Shepard, Y. Hou, and J. R. Lhota, "Review and comparison of shearography and pulsed thermography for adhesive bond evaluation," *Optical Engineering*, vol. 46, no. 5, p. 051007, 2007. [Online]. Available at <http://link.aip.org/link/?JOE/46/051007/1>
- [9] J. A. Schroeder, T. Ahmed, B. Chaudhry, and S. Shepard, "Nondestructive testing of structural composites and adhesively bonded composite joints: pulsed thermography," *Composites Part A: Applied Science and Manufacturing*, vol. 33, no. 11, pp. 1511–1517, 2002. [Online]. Available at <http://www.sciencedirect.com/science/article/B6TWN-47YXF1T-6/2/bc71b40b35ddc50e5ea83126db42480d>
- [10] D. A. Scott, M. Dennis, N. Khandekar, J. Keeney, D. Carson, and L. S. Dodd, "An Egyptian cartonnage of the Graeco-Roman period: examination and discoveries," *Studies in Conservation*, vol. 48, no. 1, pp. 41–56, 2003.
- [11] D. Bomford, *Art in the Making: Underdrawings in Renaissance Paintings*, ser. National Gallery London Publications. New Haven: Yale University Press, 2002.
- [12] B. Berrie, E. R. de la Rie, R. Hoffman, J. Tomlinson, T. Wiesel, and J. Winter, *Scientific Examination of Art: Modern Techniques in Conservation and Analysis*. Washington DC: National Academy of Sciences, 2002.
- [13] L. MacDonald, *Digital Heritage: Applying Digital Imaging to Cultural Heritage*. Amsterdam: Elsevier, 2006.

- [14] C. M. Falco, "High resolution digital camera for infrared reflectography," *Review of Scientific Instruments*, vol. 80, no. 7, pp. 071 301–071 309, Jul. 2009.
- [15] A. H. Bendada, S. Sfarra, D. Ambrosini, D. Paoletti, C. Ibarra-Castanedo, and X. Maldague, "Active thermography data processing for the NDT&E of frescoes," in *Quantitative Infrared Thermography 2010*, Quebec City, Canada, July 2010, pp. 961–966.
- [16] G. M. Carlomagno and C. Meola, "Comparison between thermographic techniques for frescoes NDT," *NDT & E International*, vol. 35, no. 8, pp. 559–565, 2002.
- [17] K. Blessley, C. Young, J. Nunn, J. Coddington, and S. M. Shepard, "The feasibility of flash thermography for the examination and conservation of works of art," *Studies in Conservation*, vol. 55, pp. 107–120, 2010.
- [18] D. Gavrilov, C. Ibarra-Castanedo, E. Maeva, O. Grube, X. Maldague, and R. Maev, "Infrared methods in noninvasive inspection of artwork," 9th International Conference on Non-destructive Testing of Art, 2008.
- [19] C. Ibarra-Castanedo, S. Sfarra, D. Ambrosini, D. Paoletti, A. Bendada, and X. Maldague, "Subsurface defect characterization in artworks by quantitative PPT," *Quantitative Infrared Thermography*, 2008.
- [20] D. Ambrosini, C. Daffara, R. D. Biase, D. Paoletti, L. Pezzati, R. Bellucci, and F. Bettini, "Integrated reflectography and thermography for wooden paintings diagnostics," *Journal of Cultural Heritage*, vol. 11, no. 2, pp. 196–204, 2010.
- [21] M. Lettner and R. Sablatnig, "Multispectral imaging for analyzing ancient manuscripts," 17th European Signal Processing Conference, August 2009.
- [22] K. Rapantzikos and C. Balas, "Hyperspectral imaging: potential in nondestructive analysis of palimpsests," in *IEEE International Conference on Image Processing*, vol. 2, Sept. 2005, pp. II–618–21.
- [23] G. A. Ware, D. M. Chabries, and R. W. Christiansen, "Multispectral document enhancement: Ancient carbonized scrolls,"

IEEE Proceedings on Geoscience and Remote Sensing Symposium, pp. 2486–2488, 2000.

[24] R. J. Easton, K. Knox, and W. Christens-Barry, “Multispectral imaging of the Archimedes palimpsest,” in *Applied Imagery Pattern Recognition Workshop*, October 2003, pp. 111–116.

[25] D. Obbink, “A new Archilochus poem,” *Zeitschrift für Papyrologie und Epigraphik*, vol. 156, pp. 1–9, 2006.

[26] B. Zuckerman, “Bringing the Dead Sea scrolls back to life: A new evaluation of photographic and electric imaging of the Dead Sea scrolls,” *Dead Sea Discoveries*, vol. 3, no. 2, pp. 178–207, 1996.

[27] G. Bearman, B. Zuckerman, K. Zuckerman, and J. Chiu, “Multi-spectral digital imaging of Dead Sea scrolls and other ancient documents,” *Jet Propulsion Laboratory, Tech. Rep.*, 1993.

[28] J. Monaghan, G. Ware, J. Pohl, and S. Houston, “A codex imaging project at the Bodleian library: the recovery of lost mixtec writing,” *Foundation for the Advancement of Meso-American Studies*, 2004.

[29] U. Bergmann and K. T. Knox, “Pseudo-color enhanced X-ray fluorescence imaging of the Archimedes Palimpsest,” *Proceedings Document Recognition and Retrieval*, vol. XVI, pp. 1–10, January 2009.

[30] X. P. Maldague, *Theory and Practice of Infrared Technology for Nondestructive Testing*. Hoboken: John Wiley & Sons, Inc., 2001.

[31] S. M. Shepard, “Understanding flash thermography,” *Materials Evaluation*, pp. 460–464, May 2006.

[32] H. Carslaw and J. Jaeger, *Conduction of Heat in Solids*, 2nd ed. Oxford: Clarendon Press, 1986.

[33] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Upper Saddle River: Pearson Education, Inc., 2008.

[34] S. M. Shepard, “Temporal noise reduction, compression and analysis of thermographic image data sequences,” Patent, 02 2003, uS 6516084. [Online]. Available at <http://www.patentlens.net/patentlens/patent/US 6516084/en/>

[35] S. M. Shepard, J. R. Lhota, B. A. Rubadeux, D. Wang, and T. Ahmed, "Reconstruction and enhancement of active thermographic image sequences," *Optical Engineering*, vol. 42, no. 5, May 2003.

[36] J. Ramirez-Granados, G. Paez, and M. Strojnik, "Reconstruction and analysis of pulsed thermographic sequences for nondestructive testing of layered materials," *Applied Optics*, vol. 49, no. 9, March 2010.

[37] K. Wan and Q. Ma, "Data-fitting reconstruction for defect inspection of airplane aluminum structure in infrared thermographic NDT," *ICIEA 2009, 4th IEEE Conference on Industrial Electronics and Applications*, 2009.May 2009, pp. 691–696.

[38] M. Pilla, M. Klein, X. Maldague, and A. Salerno, "New absolute contrast for pulsed thermography," *Proceedings of Quantitative Flash Thermography*, pp. 53–58, 2002.

[39] D. Gonzalez, C. Ibarra-Castanedo, M. Pilla, M. Klein, J. Lopez-Higuera, and X. Maldague, "Automatic interpolated differentiated absolute contrast algorithm for the analysis of pulsed thermographic sequences," in *Quantitative Infrared Thermography 7*, D. Balageas, J. Buchlin, G. Carlomagno, and G. Busse, Eds., Bruxelles, Belgique, July 2004, pp. H.16.1–H.16.6.

[40] H. Benitez, C. Ibarra-Castanedo, A. Bendada, X. Maldague, H. Loaiza, and E. Caicedo, "Definition of a new thermal contrast and pulse correction for defect quantification in pulsed thermography," *Infrared Physics & Technology*, vol. 51, pp. 160–167, 2008.

[41] C. Ibarra-Castanedo, H. Benitez, X. Maldague, and A. Bendada, "Review of thermal-contrast-based signal processing techniques for the nondestructive testing and evaluation of materials by infrared thermography," *International Workshop on Imaging Nondestructive Evaluation*, 2007.

[42] S. M. Shepard. (2007, October) Flash thermography of aerospace composites. [Online]. Available at <http://www.ndt.net/article/panndt2007/papers/132.pdf>.

[43] M. A. Omar and Y. Zhou, "A quantitative review of three flash thermography processing routines," *Infrared Physics & Technology*, vol. 51, no. 4, pp. 300–306, 2008. [Online]. Available at

<http://www.sciencedirect.com/science/article/B6TJ9-4R0643G-2/2/1012d8792b0b432e3dfd2514330f125c>.

[44] J. G. Sun, "Analysis of pulsed thermography methods for defect depth prediction," *Journal of Heat Transfer*, vol. 128, no. 4, pp. 329–338, 2006.

[45] N. Rajic, "Principal component thermography for flaw contrast enhancement and flaw depth characterisation in composite structures," *Composite Structures*, vol. 58, no. 4, pp. 521–528, 2002. [Online]. Available at <http://www.sciencedirect.com/science/article/B6TWP-472841H-9/2/4c29e1491d0beb293d9a3811ed9498b8>

[46] —, "Principal component thermography," Defense Science and Technology Organisation Victoria (Australia) Aeronautical and Maritime Research Laboratory, Tech. Rep. 012-294, April 2002.

[47] W. Swiderski, "The characterization of defects in multi-layered composite materials by thermal tomography methods," in *Proceedings of the Tenth Annual Conference of the Materials Research Society of Serbia*, vol. 115, no. 4, 2009, pp. 800–804.

[48] H. Ringermacher, D. Mayton, D. Howard, and B. Cassenti, "Towards a flat-bottom hole standard for thermal imaging," *Review of Progress in Quantitative Nondestructive Evaluation*, vol. 17, pp. 425–9, 1998.

[49] V. Vavilov, D. Nesteruk, V. Shiryayev, A. Ivanov, and W. Swiderski, "Thermal (infrared) tomography: terminology, principal procedures, and application to nondestructive testing of composite materials," *Russian Journal of Nondestructive Testing*, vol. 46, no. 3, pp. 151–161, 2010.

[50] V. P. Vavilov, "Dynamic thermal tomography: perspective field of thermal NDT," S. A. Semanovich, Ed., vol. 1313, no. 1. SPIE, 1990, pp. 178–182. [Online]. Available at <http://link.aip.org/link/?PSI/1313/178/1>.

[51] M. Omar, M. Hassan, K. Saito, and R. Alloo, "IR self-referencing thermography for detection of in-depth defects," *Infrared Physics & Technology*, vol. 46, pp. 283–289, 2005.

[52] G. Busse, D. Wu, and W. Karpen, "Thermal wave imaging with

phase sensitive modulated thermography,” *Journal of Applied Physics*, vol. 71, no. 8, pp. 3962–3965, April 1992.

[53] X. Maldague and S. Marinetti, “Pulse phase infrared thermography,” *Journal of Applied Physics*, vol. 79, no. 5, pp. 2694–2698, March 1996.

[54] X. Maldague, F. Galmiche, and A. Ziadi, “Advances in pulsed phase thermography,” *Infrared Physics & Technology*, vol. 43, pp. 175–181, 2002.

[55] C. Ibarra-Castanedo and X. Maldague, “Defect depth retrieval from pulsed phase thermographic data on plexiglas and aluminum samples,” *SPIE Proc. Thermosense XXVI*, vol. 5405, 2004.

[56] D. Gonzalez, C. Ibarra-Castanedo, F. Madruga, and X. Maldague, “Differentiated absolute phase contrast algorithm for the analysis of pulsed thermographic sequences,” *Infrared Physics & Technology*, vol. 48, pp. 16–21, 2006.

[57] M. Klein, A. Bendada, C. Ibarra-Castanedo, and X. Maldague, “A hybrid pulsed thermography processing technique for the depth estimation of subsurface defects combining TSR and PPT,” *10th International Conference on Quantitative Infrared Thermography*, 2010.

[58] B. Sun, Q. Ma, and H. Zhao, “Fitting-correlation analysis of pulsed thermographic sequence data,” in *Proceedings of the 2007 IEEE International Conference on Mechatronics and Automation*, August 2007.

[59] M. Klein, C. Ibarra-Castanedo, A. H. Bendada, and X. Maldague, “Thermographic signal processing through correlation operators in pulsed thermography,” in *Thermosense XXX, SPIE Defense and Security Symposium*, V. P. Vavilov and D. D. Burleigh, Eds., vol. 6939, Orlando, Florida, USA, Vol. 6939, March 16–20 2008, p. 693915.

[60] D. de Cogan, W. J. O’Connor, and S. Pulko, *Transmission Line Matrix in Computational Mechanics*. Boca Raton: Taylor and Francis, 2006.

[61] D. Gonzalez, C. Ibarra-Castanedo, J. Lopez-Higuera, and X.

Maldague, "New algorithm based on the Hough transform for the analysis of pulsed thermographic sequences," *Nondestructive Testing and Evaluation International*, vol. 39, pp. 617–621, 2006.

[62] C. Ibarra-Castanedo, D. Gonzalez, F. Galmiche, X. Maldague, and A. H. Bendada, "Discrete signal transforms as a tool for processing and analyzing pulsed thermographic data," in *SPIE - The International Society for Optical Engineering, Thermosense XXVIII*, vol. 6205, Orlando, FL, 2006, pp. 620 514–1–620 514–12.

[63] C. A. Larsen, "Document flash thermography," Master's thesis, Utah State University, 2011.

Surface Geometry and Heat Flux Effect on Thin Wire Nucleate Pool Boiling of Subcooled Water in Microgravity

Troy Munro, Andrew Fassmann, and Heng Ban
Utah State University

Abstract

In the summer of 2010, a nucleate pool boiling experiment was flown on NASA's Weightless Wonder to study nucleate boiling heat transfer in microgravity. The motivation for this research was to understand the effects of surface geometry and heat flux applied to a thin wire heater. This will further the understanding of the fundamental behaviors of boiling onset, steady state heat transfer, and bubble dynamics with respect to nucleate boiling, with the goal of creating efficient thermal management systems for space applications. Using three thin platinum wire geometries and five different power levels, subcooled water was boiled over a period of approximately 25 seconds during 15 parabolic arcs to simulate microgravity. To represent the trends in bubble behavior across hundreds of frames of video in a single graph, a new method, named relative bubble area analysis, was introduced and used to analyze the results of the experiment. It was determined that the efficiency of steady state heat transfer via nucleate boiling in microgravity is comparable to steady state heat transfer in terrestrial applications.

The three-wire geometry reduced the heat flux necessary to initiate boiling. Bubble dynamics show a transition from isolated bubbles to jets of small bubbles as heat flux increases. This can be confirmed both visually and with relative bubble area analysis. The implications of this research are that sustained convective heat transfer with subcooled water is possible in microgravity. A three-wire surface geometry was shown to initiate boiling at lower heat fluxes, which would provide minimal super heating of the surface, from lack of convection, before boiling heat transfer could begin.

I. Introduction

Nucleate boiling is used in a variety of terrestrial heat transfer applications because of its associated high heat transfer rates. The phase change process and the generation of highly dynamic, buoyancy-based, natural convection provide the ability to transfer more heat from smaller surface areas in relation to other forms of heat transfer. Currently, large-scale power production utilizes boiling to transfer heat from an energy source to an electricity-generating turbine. The use of boiling to cool electronic components is also becoming increasingly popular as electronics become denser and more capable; however, the current limited and empirically based understanding of nucleate boiling dynamics in the absence of the dominant buoyant force prevents the development of systems with predictable boiling dynamics for space applications. Previous microgravity boiling research has produced apparently contradictory conclusions on boiling dynamics because of the use of a variety of working fluids, surface geometries, levels of subcooling, and heat flux. The main findings of these microgravity experiments are summarized below.

A. Onset of Nucleate Boiling

The transition from natural convection (1-g) or conduction (0-g) to nucleate boiling is often studied to demonstrate the benefits of operating in the nucleate boiling regime. The dominance of conductive heat transfer in microgravity environments results in high surface temperatures according to Fourier's law until the onset of boiling is reached. It is therefore desirable for a microgravity thermal management system to operate in the nucleate boiling regime. The relationships among working fluid, heat flux, surface geometry, gravity, and onset temperature are observed to predict when this transition will occur for various combinations of these parameters and what type of heat transfer conditions will exist.

Straub [1] studied the heat transfer efficiencies associated to various heat flux densities while boiling subcooled degassed R113 (Freon R113) with a 0.2-mm-diameter platinum wire aboard the TEXUS 3b sounding rocket. Under the lowest heat flux (17 kW/m^2), the 0-g experiment boiled after reaching an onset temperature of $\Delta T_{\text{sat,wire}} = T_{\text{wire}} - T_{\text{sat}} = 34 \text{ K}$ and then dropped to $\Delta T_{\text{sat,wire}} = 18 \text{ K}$ after boiling was initiated.¹ For the 1-g experiment under the same heat flux, natural, buoyancy-based convection adequately cooled the wire, preventing the transition to nucleate boiling. At the median heat flux density (39 kW/m^2), the 0-g experiment continued to boil with a 4-K wire temperature increase from the previous heat flux level, resulting in a final $\Delta T_{\text{sat,wire}} = 22 \text{ K}$ after the wire superheated. The 1-g experiment still did not boil, but the $\Delta T_{\text{sat,wire}}$ increased to 30 K, a 20-K increase from the previous level. The final heat flux density (77 kW/m^2) produced boiling in both the 0-g and 1-g experiments. The 0-g experiment, however, showed a 6% increase in heat transfer coefficient when compared with the 1-g experiment. This experiment showed that at certain heat flux densities, 0-g heat transfer can be more effective than the 1-g equivalent; however, it never addressed whether this apparent increase was within the uncertainty of the system.

Zell et al. [2] also performed an experiment on the TEXUS 5 sounding rocket. This experiment also boiled subcooled degassed R113 with a 0.2-mm-diameter platinum wire. The onset superheated temperatures of the wires ($\Delta T_{\text{sat,wire}}$) varied heavily with gravity. The 0-g onset temperature was 25% more than the 1-g equivalent. This onset temperature difference was conjectured to be due to the fact that the fluid in proximity to the wire in the 0-g experiment was quite transient in its temperature prior to the onset of nucleate boiling (ONB). The 1-g experiment, however, had reached a steady state wire temperature because of the effects of natural convection well before boiling.

Wan and Zhao [3] used a temperature-controlled pool boiling device to boil subcooled, degassed R113 with a 60- μm -diameter platinum wire. The ONB was noted by explosive boiling, which occurred as soon

¹ Nomenclature

$h_{\mu\text{g}}$ = convection heat transfer coefficient in microgravity period

$h_{1\text{g}}$ = convection heat transfer coefficient in 1-g period

T_{wire} = average wire temperature

T_{fluid} = bulk temperature of fluid

T_{sat} = saturation temperature of fluid

$\Delta T_{\text{sat,wire}}$ = average wire temperature above fluid saturation temperature

ΔT_{sub} = subcooled temperature (bulk temperature of fluid below saturation temperature)

ΔV_{wire} = voltage difference across wire

as the heater temperature (T_{wire}) was increased to a higher set-point. This temperature set-point was the same for 1-g and 0-g, leading Wan and Zhao to conclude that the onset of boiling temperature is weakly dependent, or independent, of gravity.

These previous studies, except for that by Wan and Zhao, show that the ONB in microgravity is distinctly different than ONB in 1-g and often results in an enhancement to the heat transfer; however, they fail to specify the conditions and system parameters that determine the ONB. The current experiment aims to study the effects of wire geometry and power input on the onset of boiling by observing the heat flux required to transition from natural convection to nucleate boiling in 0-g and 1-g; the decrease in average wire temperature (T_{wire}) after the transition into boiling; and its associated explosion of bubbles along the heating element.

B. Heat Flux Effect on Steady State Heat Transfer

Once the threshold heat flux input is provided to initiate boiling, the efficiency of heat transfer can vary with the magnitude of the input heat flux and surface geometry, and is quantified in the heat transfer coefficient, h , given by Newton's law of cooling:

$$q'' = h(T_{\text{wire}} - T_{\text{fluid}}) \quad (1)$$

where q'' is the heat per unit area (heat flux), T_{wire} is the average temperature of the wire, and T_{fluid} is the bulk temperature of the fluid.

As stated before, Straub [1] discovered that under a given heat flux, where both 0-g and 1-g experiments boiled, the 0-g wire temperature (T_{wire}) was actually lower than that of the 1-g. This led Straub to conclude that the 0-g steady state boiling was 6% more efficient than 1-g boiling. This is highly unexpected since the 0-g experiment is missing the seemingly major driving force of buoyancy.

In another experiment by Straub [4, 5], the enhancement of heat transfer in 0-g appears to be a function of heat flux and wire diameter. In this experiment, two platinum wires of 0.05 and 0.2 mm were used to boil saturated R134a (Freon R134a). The 0.2-mm wire showed heat transfer enhancements of up to 10% for lower power levels, which decreased with increasing heat flux. The 0.05-mm wire showed a nearly constant reduction of about 10% and eventually burned out. Straub's [4] results also showed that 0-g boiling is either slightly enhanced or the same as a 1-g experiment for a given wire superheat temperature ($\Delta T_{\text{sat, wire}}$), but the efficiency did not seem to relate to power input.

Although Straub studied the effect of wire diameter and gravity level on steady state heat transfers with a set heat flux, there is still a

need to confirm these results across different operating conditions. The current experiment seeks to resolve these issues by studying the effects of heat flux on the steady state heat transfer efficiency by observing the heat transfer coefficient for various power levels and surface geometries during steady state boiling of subcooled water.

C. The Effects of Surface Geometry on Heat Transfer

The surface characteristics of the heating element affect bubble generation and departure dynamics, which, in turn, affect the heat transfer coefficient. As previously mentioned, Straub [4] showed that a smaller diameter wire was constantly more efficient at dissipating heat in 0-g than in 1-g, but a larger diameter wire was less able to withstand higher heat fluxes without wire burnout.

Fukada et al. [6] studied the effects of calcium carbonate fouling on platinum wires of various diameters from 0.01 mm to 0.2 mm. Saturated water was boiled in both 0-g and 1-g with bare wires and wires with a coating of calcium carbonate. The scaling on the wires created large, discrete nucleation sites, which prevented the coalescence of bubbles widely seen on the single wire. The bubble departure diameters from the scale wire were also relatively smaller. The bubble coalescence on the bare wires eventually led to the development of large engulfing bubbles, which resulted in burnout. Therefore, critical heat flux was effectively increased because of the scale on the wires.

Chyu and Mghamis [7] investigated the heat transfer enhancement achieved by connecting two hollow, stainless steel tubes in line contact. They determined that the restricted region between the tubes had low wall super heats and provided a restricted geometry that was favorable to vapor bubble formation. Because of the operation of this experiment in a terrestrial laboratory, buoyancy effects prevented the heat transfer enhancement from existing anywhere except on the upward face of the twisted tubes. A microgravity environment could allow that enhancement to be experienced by the entire wire surface.

Past research has concluded that surface geometry plays a role in boiling dynamics but has not answered the questions about specific surface characteristic effects on microgravity boiling. The current experiment seeks to fill that gap by testing the effect of two unique wire geometries (a three-wire and a four-wire twist) and by specific surface features, including a notch on the heating element.

This study aims to add to the understanding of 0-g boiling dynamics by examining the effects of heat flux, surface geometry, and gravity on the boiling of highly subcooled water by using thin wire heating

elements focusing mainly on the transition into boiling, steady state heat transfer coefficients, and bubble dynamics.

II. Experiment

A. Boiling Chambers and Heating Elements

Figure 1 shows a schematic diagram of the fluid chambers used to boil the heavily subcooled ($\Delta T_{\text{sub}} = T_{\text{fluid}} - T_{\text{sat}} = 66$ K) deionized, degassed water. Three types of 1-cm-long platinum wire heating elements were used: a single 0.005-in (0.13-mm) diameter wire, a twist of three 0.003-in (0.076-mm) diameter wires, and a twist of four 0.002-in (0.051-mm) diameter wires. Using these configurations allowed for the three distinct cross-sections shown in Figure 1c. The fluid cells also contained four type T thermocouples (TC 1-4) at varying radial distances from the heating element as well as wire probes to measure the voltage drop across the heating element (Fig. 1b).

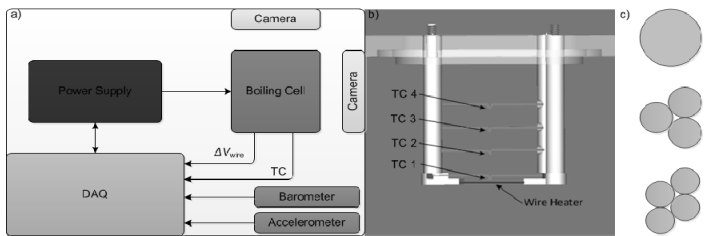


Figure 1. a) System schematic; b) boiling cell layout; c) heating element cross-sectional geometries

B. Electronics

Three power supplies provided constant current sources to each of the three heating elements. The output current was measured and recorded by a National Instruments CompactRio data acquisition (DAQ) system (Fig. 1a). The DAQ also measured the voltage drop across the heating elements (ΔV_{wire}), the voltage output from the thermocouples, the outputs from an accelerometer and a barometer (Fig. 1a), and the digital signals used to operate the experiment. Measurements of the wire voltage were effectively taken at approximately 10,000 Hz, and the mean and standard deviation were recorded at 100 Hz. The thermocouple measurements were taken at 75 Hz. Visual recordings of the experiment were taken by two Kodak Zi8 HD video cameras located

90° relative to each other to give a three-dimensional view of each cell. Each camera had a 1920×1080 -pixel resolution at 30 frames per second. A pixel resolution of $14 \mu\text{m} \times 14 \mu\text{m}$ was achieved through the use of $4\times$ magnifying lens.

C. Procedure

The experiment was performed on Zero-G's Boeing 727 micro-gravity simulator as part of NASA's Reduced Gravity Flight Program. The airplane flew 30 parabolic trajectories comprising an ~25-sec, 0-g free-fall followed by a 40-sec, 2-g climb. Each of the three different wire geometries boiled water with 5 constant power levels (Table 1).

Table 1. Average Input Current, Power, and Heat Flux Supplied to Experiment				
Geometry	“Power Level”	Input Current (± 0.01) (A)	Average Power (± 0.02) (W)	Average Heat Flux (± 50) (kW/m²)
Single-Wire	5	2.13	2.48	599
	6	2.37	3.11	750
	7	2.60	3.36	786
	8	2.81	4.40	1031
	9	3.01	4.93	1168
Three-Wire	5	2.20	2.35	397
	6	2.46	3.17	503
	7	2.71	3.32	559
	8	2.91	4.41	670
	9	3.12	4.48	777
Four-Wire	5	1.69	2.47	512
	6	1.91	3.04	639
	7	2.09	3.66	689
	8	2.26	4.44	930
	9	2.42	4.70	992

III. Results and Discussion

The wire temperature (T_{wire}) as a function of time of a single wire experiment is shown in Figure 2, with $t=0$ representing the moment that power was supplied to the wire. In analyzing nucleate pool boiling, it is important to note that there are four distinct regions that occurred during the boiling process of the current experiment. These regions are superheating, boiling onset, a transition state, and steady state boiling.

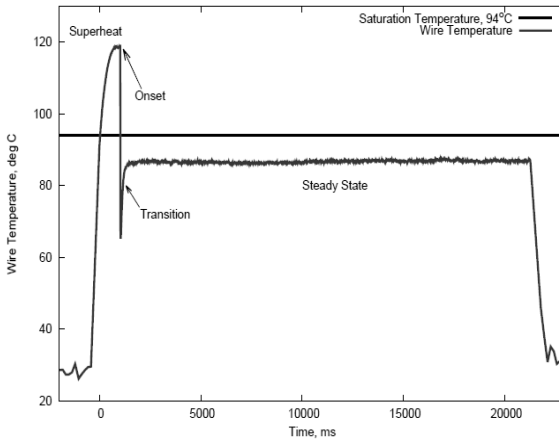


Figure 2. Wire temperature of single wire showing superheat, onset, transition, and steady state regions.

The superheating region is noted at lower heat flux levels as having extremely high wire temperatures (T_{wire}), resulting in a high $\Delta T_{\text{sat,wire}}$. After the necessary heat flux is dissipated by the wire, boiling begins. This onset condition is characterized by a rapid cooling of the wire, and at low heat fluxes, an explosion, as bubbles are seeded for the first time. This cooling behavior is similar to the observations of Zell et al. [2], but a greater cooling effect was observed in the current experiment because of the larger degree of subcooling. The transition state is characterized by a large amount of nucleated bubbles. These bubbles grow and depart until equilibrium of bubble nucleation, growth, and departure is reached. The regions of greatest interest—boiling onset, boiling steady state, and steady state transition to jets—will be analyzed below.

A. Relative Bubble Area Analysis

Because of the large amount of visual data collected, a computerized relative bubble area analysis was developed in an effort to quantify the change of bubble area from one picture frame to next. In general, this method determines and measures the area of the picture covered by bubbles. This allows for bubble nucleation, growth, and departure to be summarized for an entire frame.

After the images were stabilized, one image was selected as a base from which to measure changed pixels. Each successive image was then taken and, effectively, had the base image subtracted from it. The effect was the ability to count the number of pixels changed over a threshold value to minimize false positives while still including behavior of the entire wire. The number of changed pixels was then plotted against the frame number, or effectively, changed pixels over time.

Figure 3 shows the change in area occupied by bubbles as a function of elapsed heater time of a single wire geometry. Images in Figure 3a-d are associated with the indicated point in time and represent the viewing area tracked in the relative bubble area analysis after the images have been stabilized. These images show an evolution as bubbles nucleate and transition to steady state boiling. Even though some noise can be seen over the plotted graph the trends indicated on the graph give accurate summaries of the bubble dynamics observed during the 30 seconds of video. The comparison of bubble area for different experimental conditions is discussed in the sections below, further illustrating the ability of this method to quantify trends present across multiple video frames.

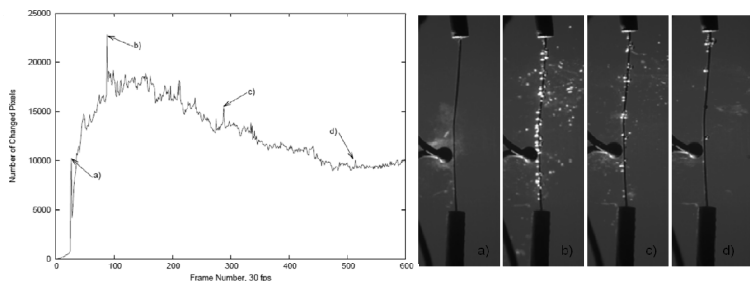


Figure 3. Relative bubble area analysis plot of single-wire geometry at power level seven (786 kW/m^2), showing corresponding images.

B. ONB

The ONB can be easily seen in Figure 4, which depicts the three different wire geometries as they transition from conduction to nucleate boiling in 0-g. The onset of boiling can be seen as a large drop in wire surface temperature (T_{wire}). This temperature drop is caused by sudden increase in the convection coefficient as boiling begins with an explosion of bubbles, as previously observed by Wan and Zhao [3]. This bubble explosion is seen as a spike in the number of changed pixels as evident in the first peak of Figure 3 (and Fig. 3a) and is one of the key features of boiling onset. After the initial explosion, bubbles begin to form along the majority of the wire (Fig. 3b). Such explosions occur with all three wire geometries. It seems that this explosion is the conduit in which bubbles are seeded onto the nucleation site that allows the bubble growth to follow. These explosions are observed mostly at lower power levels that still produce boiling, but at higher power levels, this pattern may not be able to be seen.

Superheating before boiling begins is caused by the lack of natural convection in microgravity. Once boiling begins, more heat can be transferred away from the wire and thus causes the decrease in wire temperature. The values used to determine onset temperature, heat flux, power, and heat transfer coefficient were taken at the last point before the drop in temperature. If this spike was not observed, as is the case in many of the 1-g experiments, the ONB was determined to be the first data point for the first power level where boiling was present. In comparing the values of boiling onset heat flux, we see that the three- and four-wire geometries lower the heat flux required in boiling. The single wire required a heat flux of 823 kW/m^2 , whereas the four-wire geometry required a value of 586 kW/m^2 . The three-wire geometry's onset condition was found to be a range between 396 kW/m^2 and 519 kW/m^2 . The reason for this range is that both the single-wire and four-wire geometries reached a steady state superheating condition before transitioning into boiling (Fig. 4). The three-wire geometry did not boil in the lowest power and did not have a steady state superheating region in the next power level up (Table 1). This indicated that the transition into boiling occurred between these two power levels or heat fluxes.

Even though the three-wire geometry has a range for its onset condition, the entire range remains below that seen in the single- and four-wire geometries. This geometry can thus effectively lower the onset heat flux for boiling. It is proposed that the reason for this lowering of onset heat flux is due to the area of enhanced temperatures in the

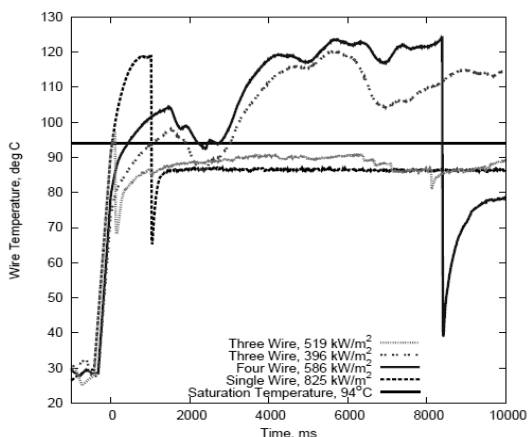


Figure 4. Wire temperatures at onset of boiling.

cavity between the three wires, which causes local superheating. This area would allow for an even greater amount of local superheating to occur, which would be more optimal for seed bubble production. These bubbles would then be allowed to seep to the outside surface of the wire and continue to dissipate heat from the wire. The four-wire geometry also has this superheated region, but it seems that it is not as effective, because the two-cavity situation in the four-wire case needs more power to generate the same effect with its larger ratio of inner and outer areas (1:4) compared with the three-wire case (1:6). The bubble dynamics seem to suggest that it more closely approximates the single wire. The ratio of area within the superheated region to area outside of the superheated region may be related to the ability for the combination of wires to approximate the single-wire geometry (1:1).

C. Steady State Nucleate Boiling

After the initial onset of boiling, the system reached a steady state region characterized by a nearly constant wire temperature (T_{wire}). This behavior can be seen in the steady state region shown in Figure 2. The slight dips in wire temperature seen on the three-wire geometry (Fig. 4) correspond to departures of large bubbles from the heater, resulting in a lower wire temperature as the departing bubble is replaced by the sub-cooled water. It was also observed that as the heat flux dissipated by the wire increased, the steady state temperature of the wire also increased, similar to the trend observed by Straub [1]. The current ex-

periment's large amount of subcooling, however, dropped the average wire temperature below that of saturation ($\Delta T_{\text{sat wire}} < 0$).

To compare the efficiency of boiling in microgravity to terrestrial tests, the same conditions found on the Boeing 727 were reproduced in a 1-g environment. The values computed for heat flux and the heat transfer coefficient were averaged over this steady state region after boiling had been initiated, and the heat transfer coefficient, h , was calculated by Newton's law of cooling. Figure 5 shows the ratio of the heat transfer coefficient in microgravity to heat transfer coefficient in 1-g versus the power at which the heat transfer coefficient was achieved.

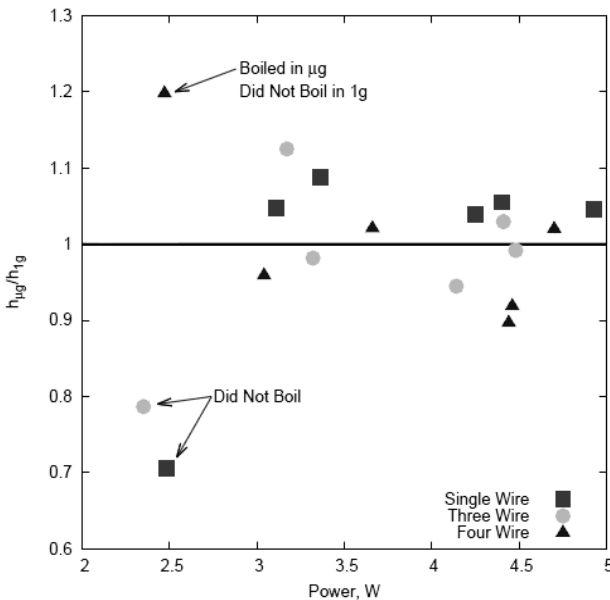


Figure 5. Comparison of heat transfer coefficient in microgravity and on earth for different power inputs.

When the ratio is greater than 1, heat transfer is more efficient in microgravity; when the value is less than 1 it is less efficient. This plot shows that within the 10% uncertainty in heat transfer coefficient values (due to measurement uncertainties), gravity plays an independent role in influencing the effectiveness of heat transfer. Two exceptions to this observation exist as special cases. The first is the four-wire geome-

try at 2.5 W, when the system boiled in microgravity but did not boil on Earth. This resulted in a much higher heat transfer coefficient in microgravity, thus showing the benefit of using boiling over natural convection. The other cases (single- and three-wire geometries at 2.5 W) are characterized by a lack of boiling in both 1-g and 0-g tests. This resulted in conduction dominant heat transfer in microgravity compared with the natural convection dominant heat transfer in the terrestrial test.

D. Bubble Dynamics

Although steady state boiling characteristics are independent of gravity, the bubble dynamics are different. As a general rule, larger bubbles are able to form on the wire in microgravity because of the lower frequency at which bubbles depart because of lack of buoyancy. As heat flux increases, there is a transition from the formation of isolated bubbles to the formation of bubble jets. These jets were similar in appearance to fog-like liquid–vapor jets seen by Wang et al. [10, 11], which had been observed in subcooled alcohol. This transitional behavior can be seen in Figure 6 as the relative bubble analysis shows a decrease in the number of changed pixels as time passes.

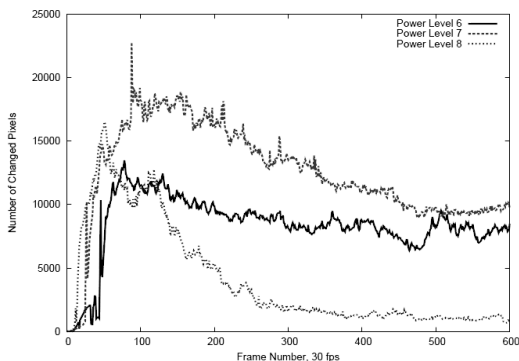


Figure 6. Relative bubble area analysis results for 3 different power levels of the single wire showing transition from isolated bubbles to jets.

The larger values represent large isolated bubbles and the smaller values represent the dominant behavior of jets. The reason for this discrepancy between changed pixels is the small size of the jets relative to the bubbles and the inability of the system lighting to show all of the

jets along the wire. Some jets are visible because they are contained in the stream of light; however, many remain invisible and can only be seen through their influence on other bubbles.

These observations suggest that in microgravity, jets provide a more efficient way for a large amount of heat to be transferred from the wire as opposed to isolated bubbles removing the heat. This is due to the formation of convection currents that have a pumping effect [12, 13] of pushing the hot liquid–vapor water away from the wire in the absence of a dominant buoyancy force. This behavior allows the temperature of thermocouple 2 to be greater than thermocouple 1 (which is closer to the heating element as seen in Fig. 1b) as heat is transferred by the convective jet in its proximity.

As the heat flux dissipated by the wire increases, the transition from isolated bubbles to jets occurs (Fig. 6). At the highest heat fluxes, jet behavior dominates bubble dynamics and there is virtually no observation of transitions from isolated bubbles occurring. This behavior is probably due to the presence of Marangoni effects in the subcooled water which is a dominant factor on the formation of these jets [13]. The geometry of the wire heater also affects the tendency of the boiling system to form jets. The bubble dynamics of the three-wire geometry have a tendency to form large, isolated bubbles that either coalesce into larger bubbles or depart. The single- and four-wire geometries had very similar behavior with the initial formation of smaller isolated bubbles and the transition to jets. This transition was more rapid for the four-wire than the single wire.

The platinum wires used in this experiment were magnified 100 \times under an optical microscope after testing had been performed with the wires to investigate the nucleation sites from which these convective jets originate. Of special interest is a single wire that experienced a jet in both microgravity and terrestrial experiments. In all configurations (microgravity in Fig. 7c, vertically oriented in Fig. 7d, and horizontally oriented in Fig. 7e), a jet formed on the same location of the wire (shown in Fig. 7a). The ability to associate the formation of this jet with a specific nucleation site on the wire helps to resolve the observation of Wang et al. [14] that these jet flows are random and unpredictable. Further work to predict the formation of these jet flows, specifically with the twist of three wires, is needed and is the subject of an ongoing microgravity nucleate boiling experiment.

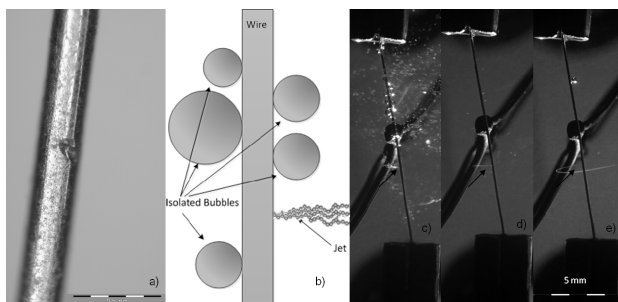


Figure 7. a) Nucleation site of jet on single wire; b) diagram of jet bubble dynamics; c) jet in microgravity; d) jet in Earth gravity (vertical orientation); e) jet in Earth gravity (horizontal orientation).

IV. Conclusions and Recommendations

On the basis of the results of this microgravity experiment, we conclude that nucleate boiling can be an efficient means of transferring heat in microgravity under certain conditions. This observation is based on the following conclusions:

- In many instances, steady-state heat transfer is independent of gravity.
- The unique twist of three wires provides a surface geometry that reduces the required heat flux for onset boiling. This behavior is favorable because more heat is transferred at smaller ΔT_{sat} .
- As heat flux increases, there is an increased tendency to form jets, which provide convective currents normally absent in microgravity.

It appears that the heat flux dissipated by the wire is the primary parameter in determining the mode of heat transfer experienced in microgravity, whether it is pure conduction, boiling with isolated bubbles, or boiling with convective jets.

Further work includes determining the onset heat flux of the three-wire geometry, determining the heat transfer performance of the wire geometry near critical heat flux, predicting jet formation and performance, and applying the results from previous experiments to design an array of microheaters to provide heat transfer enhancement by creating seed bubbles.

Acknowledgments

All the members of the USU Get Away Special (GAS) team of 2010 are acknowledged for researching, designing, constructing, and performing this experiment. Numerous organizations and individuals donated funding or equipment, including National Instruments, SpaceX, USU Space Dynamics Laboratory, American Aerospace Advisors, Rocky Mountain NASA Space Grant Consortium, USU Physics Department, USU Mechanical and Aerospace Engineering Department, USU College of Science, College of Engineering, USU Undergraduate Research Office, Associated Students of USU, Gil Moore, Scott Thomas, and Russ Laher. This study was made possible by NASA's Reduced Gravity Student Flight Opportunities Program, which provided the opportunity to perform this experiment in a microgravity environment.

Special thanks are overdue for Justin Koeln, Robert Barnett, and Dr. JR Dennison for all their help with design, analysis, and preparation of this paper, and to Dr. Jan Sojka for his work as the GAS team advisor.

References

- ¹Straub, J. "Highs and lows of 30 years research of fluid physics in microgravity, a personal memory," *Microgravity Science and Technology*, Vol. 18, No 3-4, 2006, pp. 14-20.
- ²Zell, M., Straub, J., Weinzierl, A., "Nucleate Pool Boiling in Subcooled Liquid under Microgravity—Results of TEXUS Experimental Investigations," *Proceedings of the 5th European Symposium on Material Sciences under Microgravity*, EPA SP-222, Schloss Elmau, Germany, 5-7 November 1984.
- ³Wan, S.X., Zhao, J.F., "Pool Boiling in Microgravity: Recent Results and Perspectives for the Project DEPA-SJ10," *Microgravity Science and Technology*, Vol. 20, 2008, pp. 219-224.
- ⁴Straub, J., "Bubble—Bubbles—Boiling," *Microgravity Science and Technology*, Vol. 16, No. 1, 2005, pp. 242-248.
- ⁵Straub, J., Zell, M., Vogel, B., "What We Learn From Boiling under Microgravity," *Microgravity Science and Technology*, Vol. 6, 1993, pp. 239-247.

⁶Fukada, Y., Haze, I., Osakabe, M., "The Effect of Fouling on Nucleate Pool Boiling of Small Wires," *Heat Transfer-Asian Research*, Vol. 33, No. 5, 2004, pp. 316-329.

⁷Chyu, M.C., Mghamis, A.M., "Nucleate boiling on two cylinders in line contact," *International Journal of Heat and Mass Transfer*, Vol. 34, No. 7, 1991, pp. 1783-1790.

⁸Zell, M., Straub, J., Vogel, B., "Pool Boiling Under Microgravity," *PhysicoChemical Hydrodynamics*, Vol. 11, No. 5/6, 1989, pp. 813-823.

⁹Rohsenow, W.M., "A method of correlation heat transfer data for surface boiling of liquids," *Transactions ASME. Series C, Journal of Heat Transfer*, Vol. 74, 1952, pp. 969.

¹⁰Wang, H., Peng, X.F., Wang, B.X., Lee, D.J., "Jet flow phenomena during nucleate boiling," *International Journal of Heat and Mass Transfer*, Vol. 45, 2002, pp. 1359-1363.

¹¹Wang, H., Peng, X., Garimeela, S.V., Christopher, D.M., "Microbubble return phenomena during subcooled boiling on small wires," *International Journal of Heat and Mass Transfer*, Vol. 50, 2007, pp. 163-172.

¹²Shekriladz, I.G., "Discussion: 'Dynamics of Bubble Motion and Bubble Top Jet Flows From Moving Vapor Bubbles on Microwires' (Christopher, D.M., Wang, H., and Peng, X., 2005, *Journal of Heat Transfer*, 127, pp. 1260-1268)," *Journal of Heat Transfer*, Vol. 128, 2006, pp. 1343-1344.

¹³Wang, H., Peng, X.F., Christopher, D.M., Lin, W.K., Pan, C., "Investigation of bubble-top jet flow during subcooled boiling on wires," *International Journal of Heat and Fluid Flow*, Vol. 26, 2005, pp. 485-494.

¹⁴Wang, H., Peng, X.F., Wang, B.X., Lee, D.J., "Bubble sweeping and jet flows during nucleate boiling of subcooled liquids," *International Journal of Heat and Mass Transfer*, Vol. 46, 2003, pp. 863-869.

A Comparative Analysis of Prioritizing Professional Behavior between Undergraduate Students of Exercise Science at Utah Valley University and Future Employers

**Bret H. Boyer, Paige Denney, M. Bohne, and Jason
Slack**

Utah Valley University

Abstract

Students often put a lot of time and energy in professional behaviors that are different than what graduate programs and employers are actually seeking. The purpose of this study was to compare the perceived priority of professional behaviors between general Utah Valley University students, Exercise Science students, and employers in the realm of exercise science. A total of 572 students were surveyed to rank a randomized list of professional behaviors that were statistically compared with those of potential employers of this same student population. Analysis reveals that other than the #1 ranking, the classes differed from each other in ranking professional behavior and they differed from employers. Having identified these differences, we might consider designing academic programs to bring the students more in line with

employers' priorities and doing it earlier in the students' academic careers.

Introduction

Many students report that they are seniors before they are aware of the professional behaviors that employers are seeking. They then discover that their own priority of these same professional behaviors is not the in line with the employers' priorities, and thus they have been putting too much their energy in the wrong areas. While this is a common complaint from students, we were not able to find any research to statistically confirm whether there are actually any differences (1).

Statement of the Problem

The purpose of this study was to compare the perceived priority of professional behaviors between general Utah Valley University (UVU) students, exercise science students at UVU, and potential employers of Exercise Science graduates.

Significance of the Study

The significance of this study is severalfold. It will enable us to statistically:

1. Identify how each college class prioritizes the behaviors;
2. Compare the priorities among the different college classes;
3. Compare the priorities of GE students with those of Exercise Science students;
4. Compare the priorities of the different student populations to those of the employers.

Having made the comparisons, we can then take appropriate action, if deemed necessary, to affect an academic design to bring the students more in line with the employers' priorities, and doing it earlier in the students' academic careers.

Hypothesis

We hypothesize that there is a significant difference in prioritizing professional behaviors among between classes and between classes and employers.

Research Methods

With UVU Institutional Review Board approval, an assessment tool was designed listing 10 professional behaviors that were identified having solicited faculty, graduate programs, and employers. Some demographics such as year in school, age, program of study, career goals, and type of employment – for employers, were asked on the survey. The students and employers were instructed to rank a list of 10 professional behaviors from 1 to 10, marking 1 for most important and 10 for least important, based upon their understanding in regard to what employers are looking for upon graduation. Each number was to be used only once, and if a behavior was not understood, then they were to leave that one blank. Three different surveys were handed out in sequence, each having the behaviors listed in one of three different sequences to reduce the bias of the order the behaviors were listed on the survey tool. Those behaviors listed included:

- Work-related Experience
- Letter of Recommendation
- Charitable Service
- Attendance at Professional Conferences
- Undergraduate Research
- Work Portfolio
- Members Professional Organizations
- Grade Point Average (GPA)
- Internship
- Presenting at Conferences (oral / poster)

The tool was administered to 572 students and 31 different potential employers of exercise science majors. A total of 32 surveys were thrown out because the survey was not completed as instructed.

A Kruskal-Wallis Ranks Test and a Tukey Hsd Post Hoc Test were used to compare and statistically analyze the average perceived priority of professional behaviors between classes of students and employers.

Results

A significant difference in prioritizing professional behavior was found between each population.

Analysis reveals that other than the behavior of “work-related experience,” which averaged #1 ranking with all classes and employers, there was a significant difference in rankings between each population average. First, second, and third rankings were spread between 4 of the 10 behaviors: work-related experience, letters of recommendation, portfolio, and GPA. Juniors and employers ranked portfolio as #2. Sophomores, seniors, and GE students ranked GPA as #2. Seniors and employers ranked letter of recommendation as #3. Sophomores and GE students ranked portfolio as #3, and juniors ranked GPA as #3.

It was further revealed that the younger the class, the greater the difference from the employer and the more diversified the rankings.

Discussion

Limitations

Some of the limitations of this study include a solicited rather than standardized list of professional behaviors or even values (2). A standardized list such as those used by the Physical Therapist Centralized Application Services (PTCAS) (3) would possibly be more representative of what employers are looking for as far as demonstrated behaviors. A more equal proportion of the different populations would possibly demonstrate a more valid analysis.

Recommendations and Further Research

Clarifying these differences in perception of professionalism between these populations demonstrates a need for us to expose the students at an earlier point in their academic careers to professional behaviors and, at the same time, bring them more in line with employer expectations. This will in turn enhance our effectiveness as educators to facilitate student success, career goals, and the expectations of employers. Further research needs to be done with specific disciplines, undergraduate and graduate programs and between graduate programs and employers to further identify any discrepancies and thus inefficiencies in program designs. Groundwork would be beneficial to establish standardized behavior and values

Bibliography

1. Santasier, Anita M.; Plack, Margaret M.; Assessing professional behaviors using qualitative data analysis. *Journal Of Physical Therapy Education*. Education Section, American Physical Therapy Association.

Volume 21 No. 3, Winter 2007

2. Scarpaci, Jim; Musing on Professionalism. *Journal Of Physical Therapy Education*. Education Section, American Physical Therapy Association. Volume 21 No. 3, Winter 2007

3. PTCAS, Physical Therapist Centralized Application Services, <http://www.apta.org/>

Eat to Live, Live to Dance: Preventing Eating Disorders in Dancers

Kristy Jo Hunt

Dixie State College

ABSTRACT

Dancers often succumb to both internal and external pressures to maintain an ultra-thin appearance. Studies indicate this pressure results in 50% of dancers struggling with eating disorders such as anorexia and bulimia. Although the past few decades' research further identifies the problem, few preventative strategies have been implemented. Without proper education when challenged with inevitable negative body image, young dancers resort to their own means to change their bodies—unbalanced diets, self-starvation, and purging food right after eating—to attain their desired aesthetic appearance. I propose that a proper nutrition model be taught in all dance instructional facilities, both public and private, for dancers aged nine to eighteen years old.

The Body for Life nutrition model was introduced to society by Bill Phillips in 1999 in his book *Body for LIFE: 12 Weeks to Mental and Physical Strength*. This model was published in *The International*

Journal of Sport Nutrition and Exercise Metabolism in 1996 as a program that is a healthy, enabling method for individuals to reach their physical appearance goals. This nutrition model will provide proper nutritive eating information to aid dancers in reaching their optimal figures and body fat percentages while reducing numbers of eating disorders, disease, injury, and even mortality.

While there have been several propositions in existing academia for instructors to educate their students about healthy eating patterns and positive body image, a definitive nutrition model on which to focus has not yet been introduced. My research fills a void that will aid instructors in their efforts to prevent eating disorders in their students by fueling them with nutritive knowledge.

A few days ago, a girl with whom I dance on a college company privately approached me. She knew I constantly engage in nutritive studies and asked for help making a diet she could follow for a few weeks to lose 10 pounds for the upcoming concert. After I dug a little deeper, she admitted to me her obsession with her weight, preoccupation with food, self-loathing for her body shape, and her on-and-off bulimic practices for the past seven years. Sadly, this is an extremely common situation in the unique dance community where, instead of building self-confidence, the art form of dance becomes a vehicle for self-destruction.

From the time a young girl sets foot onto the floors of the dance studio, she must face the demons that result in many dancers struggling with disordered eating. Because of the secretive nature of the isolated dance community, recent and reliable statistics of all-genre dancers with disordered eating patterns are still undocumented; however, researchers have studied statistics on genre-specific dancers—particularly ballet dancers. Researcher Brenda Griner reported that in 1981, a study indicated that “50 percent of dancers are bulimic or anorectic at some time” (Gordon). Researchers surveyed 45 female ballet students between the ages of 12 and 21 years as well as 44 nonathletic females. The ballet dancers showed characteristics of anorexia nervosa (intense fear of gaining weight accompanied by drastically reduced caloric intake) “significantly more often than the nonathletic subjects (Braisted et al., 1985).

Documented research continues to reveal numbers that indicate there is a dire problem. Wendy Oliver (2008) of the National Dance Alliance stated that out of 107 female college dancers, 76% of those dancers agreed with the statement, “I think there are pressures to lose weight in dancing” (p. 22). In addition, Catherine Culnane indicated in her study of nutritional status among female dancers that on the Uni-

versal Dancer Profile, 76% of the surveyed dancers scored in the Body Dissatisfaction range (p. 95).

Where researchers have documented several eating disorder intervention programs in the past decade, they have addressed dancers' problems only after they were suffering (Griner, 2006, p. 45) or suggested dance companies' compliance with unrealistic suggestions. Griner's research suggested dance teams and companies not set specific weight or body fat requirements for membership" (Montanari and Zietkiewicz, 2000, p. 46). Researchers unfamiliar with the dance community may easily suggest this possibility, but as one who is heavily involved in that community, I know there will never be wide-scale compliance with this suggestion. What dancers need is a proper understanding of what healthy weight loss entails—a program that enables them to "find ways to achieve a lean body, eat healthy, and develop more positive attitudes about diet, nutrition, fitness, and their body image without incorporating health risk factors (Griner, 2006, p. 45). This could be done as Schluger (2011) reports using educational endeavors that "concentrate on the promotion of healthy lifestyles and self-image (Lindner et al. 2008, p. 127).

An ideal nutrition model for achieving a dancer's ideal body shape based on genetic predisposition while protecting her physical, emotional, and mental health exists. This model is called the Body for Life nutrition model (BFLNM). As adolescent female dancers aged nine through eighteen learn about this nutrition model from their dance instructors and educators, they will engage in healthier, more nutritious lifestyles while reaching their goals in aesthetic appearance. Although male dancers are also susceptible to the overwhelming pressures of aesthetic appearance that lead to eating disorders and though this research does, in fact, have just as much applicability to them, I will focus on the majority of dancers—females.

The Problem

To some degree or another, every dancer will feel the self-berating voices of negative body image; however, when those voices surface, dancers turn to their own understanding of weight loss. Without proper nutrition, weight loss, or body sculpting education, dancers turn to fad diets, self-starvation, compulsive exercising, and laxatives—all with the goal of changing their body composition. This disordered eating leads not to health problems such as high blood pressure, diabetes, and a decreased immune system, but even to heart problems that result in death.

As Oliver (2008) states, college dance instructor Julie Sandler (1997) wrote the following:

In more than 20 years of teaching, I have watched either the expressions of doubt, fear, despair, and self-disgust or alternatively, of contempt and self-righteous smugness in supposedly bodily superiority move across the faces of my female students when the issue of the dancer's body arose, or when they faced the studio mirror. The forms of being bodily wrong that students despair over are endless" (p. 18).

The demand for an ultra-slim body shape is very nearly an absolute requirement of the dance community, a special subculture that places emphasis on values that general cultures claim as irrational. This requirement is understood by dancers from verbal statements, weight requirements, and their own personal expectations because of the media. In a survey of more than 100 dancers, 57% "identified their teachers and audiences as stressors" (Oliver, 2008, p. 22).

Of the various genres of dancers, researchers have studied female ballet dancers the most extensively regarding eating disorders. George Balanchine—a famous teacher at the New York City Ballet—and his aesthetic preferences have been labeled as the reason for the "demand for dancers to have a very low weight" (Brooks-Gunn, 1988, p. 750) by repeatedly telling his dancers to "Eat nothing" (Oliver, 2008, p. 21). One of his dancers, Gelsey Kirkland, described in her autobiography *Dancing on my Grave* (1987) how her mentor's statements triggered her eating disorder. Although Balanchine's case is at the extreme end of the spectrum, dancers may interpret external pressure from instructors' or peers' statements even with well-intended statements and evaluations.

While discussing the topic of negative body image with dancers on companies with whom I rehearse and perform, it becomes apparent that every dancer either knows or knew someone with a current eating disorder or knows someone who has left the field because they needed to remove themselves from the negativity. One dancer, who wished to remain anonymous, said the following: "Though they don't intend it, there is pressure from instructors when they say things like, 'Hold in your gut,' or, 'Suck it in.'" Even the mirror with its ever-present feedback about the shapes and lines of a dancer's body "can become a conduit for negative self-talk" (Oliver, 2008, p. 23). Because "distorted body image begins as teens develop sexual characteristics," the combination of these factors can be the catalyst to self-destruction for a dancer (Duyff, 2006, p. 434).

Although ballet dancers are primarily recognized as the most-affected population at risk for eating disorders, studies indicate “no significant differences between modern dance and classical ballet dance majors in terms of their attitudes regarding perfectionism” (Schluger, 2011, p. 119). Research has also found that dancers (of different genres) scored higher on measures of perfectionism than non-dancers on the *Eating Disorder Inventory* and foster “high achievement standards” (Schluger, 2011, p. 119). This means that dancers are extremely demanding of themselves and are their own worst critics.

A dancer must endure hours in front of mirrors that spit back the image of her tool for the art. Not only is she clad in a skin-tight leotard, tights, or spandex shorts, but her eyes, trained to spot any misalignment or image of error in the mirror in order to “fix” it, instantly affix upon her own image. Comparisons begin between herself and the other students in the class, and from there to her understanding of what the professional world of dance expects of its dancers. As researcher Claudia Ravaldi (2006) found, if a dancer is to succeed, she must attain that image (p. 530).

While negative body image appears to be inevitable for female dancers, their resultant *actions* are due to lack of proper weight loss and maintenance education. Without proper education of a nutrition model that will provide adequate nutrition as well as chisel a lean, dancer figure, dancers will continue turning to their own understanding of weight loss—dieting.

Dieting Setbacks

As a dancer feels the need to lose weight, she will turn to her cultural connotations of what weight loss entails—dieting—the age for which is becoming younger and younger; some girls “as young as nine are on diets, and many others join them when they reach the teen years” (Oliver, 2008, p. 20). Diets are risky “especially during adolescence when teenagers’ nutrient needs for growth and energy are high” (Duyff, 2006, p.47).

The word “diet” commonly means food habitually eaten, particular selection of food designed to improve a person’s physical condition, and such a selection on the amount a person eats for reducing weight. Dancers may rely on extreme fad diets as well as their own interpretation of how they can lose weight. Nonetheless, when dancers do not lose the weight they desire, they turn to more drastic measures that may result in anorexia, or bulimia both through vomiting, or excessive exercise to “purge” unwanted calories.

Common problems of fad diets, or even diets constructed by uneducated individuals, include the following: a nutrient deficiency of recommended daily allowances by the American Dietetic Association, long periods of hunger that lead to binge-eating triggers, recommendation of foods that are the wrong combination for proper hormone stimulation to result in weight loss, the cognitive effects of self-deprivation, and the low percentage of long-term compliance. Dancers go through the rigors of adhering to the diet, but because it is not designed correctly for healthful, proper diminishing of fat loss, especially when there is little body fat to begin with, a dancer will not, and cannot, comply. During this time of non-compliance, she may binge eat, with or without purging afterward, or gain back the weight she lost while on the diet. She will consider herself a failure and believe she must take further action to see results. After some time has passed, she will try again, only to fall into the same cycle. This is termed ‘yo-yo’ dieting. She will have taken her first steps down the road of a lifetime struggle with self-loathing. This is a road that has resulted in many dancers either leaving their studies of the art form or becoming both physically and mentally ill.

Many times, dieting involves meal skipping. The American Dietetic Association identifies that many girls and women (not just dancers) skip breakfast or lunch to reduce their daily caloric intake. In doing so, they “miss out on nutrient-rich foods of special concern: high-calcium foods such as milk, iron-fortified cereals, fiber-rich whole grains, and fruits and vegetables.” As an individual skips meals, the body reacts with a hunger trigger. “The body perceives a diet as famine and strives to protect itself from starving to death by signaling hunger” (Yannakoulia, 2002, p. 30). Hunger leads to the urge for the body to satiate itself—usually in the form of binge eating. Food binges can be the result of going to extreme measures in order to lose a few pounds.

Other frequently used means of controlling weight are over-the-counter diet pills, appetite suppressants, or laxatives. These pills work, but have “unpleasant side effects and some can be addictive, with potential harm to the heart and nervous system” (Duyff, 2006, p. 46). A quick weight-loss regimen too low in calories may interfere with a dancer’s physical performance and may cause a dancer to lose muscle—along with body fat—and deplete her stores of muscle glycogen. Weight loss may be partly water loss—a problem for athletes who need to keep adequately hydrated. Dieting tends to create more problems physically and mentally from the deprivation than the original problem of an unsatisfactory weight or body fat percentage. Therefore, when body weight or fat reduction is necessary or desired, a proper *lifestyle*—not diet—should be implemented.

Complying with any diet long term is not usually easy because diets limit the individual as to what she cannot eat, in opposition to the body's natural eating tendencies. In addition, the American Dietetic Association says that "many diets make overblown claims and tout products that are ineffective and costly." Not only are most diets "far too restrictive and difficult to follow" (Daoust and Daoust, 2001, p. 57), but they "recommend foods that stimulate the wrong hormonal reaction to burn fat and lose weight" (Griner, 2006, p. 42). Because of this improper hormonal balance, a dancer will ultimately fail in reaching her goals of weight and body fat reduction.

A proper diet should "specify portion control ... and take into special consideration [an individual's] specific requirements" (Daoust and Daoust, 2001, p. 58). For any diet or nutrition program to be successful, "it should be easy to follow and provide detailed meal plans that show exactly what to eat. It has to have appealing foods, needs to make sense, and should be one an entire family can use" (Daoust and Daoust, 2001, p. 58). Each of these recommendations is met by the BFLNM.

Educator Intervention

Eating disorder prevention programs—which focus on educating dancers about what eating disorders are and why they are a negative way to combat their negative body image—have been labeled by other studies as largely unsuccessful. "Eating disorder prevention programs exacerbate eating and weight concerns of vulnerable or high-risk individuals" (Yannakoulia, 2002, p. 30). Therefore, a different approach—one in which the dance environment becomes "supportive, not abusive, and we strive to make positive comments as well as corrections"—needs to be taken (Oliver, 2008, p. 24). This approach must, instead of focusing on eating disorder education, "concentrate on the promotion of healthy lifestyles and self-image" (Oliver, 2008, p. 24).

Those who teach, motivate, and inspire developing dancers in the studio each day are those who will make a difference in preventing the decades-old issue of dancers developing eating disorders. Dance educators should "learn to design beneficial training programs" (Griner, 2006, p. 45) and "implement educational endeavors" that address the nutritive benefits to their dance training as well as developing the ideal, lean dancer body (Schluger, 2010, p.125). Griner's (2006) study on the challenges of meeting the demand for a lean body while maintaining a nutritious diet indicate that further research remains to be done in this field, specifically: "healthy ways to lose fat when little fat exists, and ways to speed fat loss in order to improve compliance among dancer

dieters” (p. 44). She further suggests that “studio owners, company managers, physical educators, and dance team coaches educate their students by offering regular informational lectures on nutrition and dance-related health topics.” The BFLNM should play a considerable role in those lectures.

The BFLNM has been proven successful by non-dancing women over the course of the last 11 years. Their stories indicate that anybody and any *body* can lose fat by adhering to this nutritionally sound plan. While BFLNM is initially recognized for enabling overweight persons to shed excess body fat, it is also recognized for taking average bodies and transforming them into lean bodies of extremely low body fat percentages.

If each instructor of adolescent dancers ages nine through eighteen, including public and private schools, private dance studios, and dance companies, begins implementing training workshops on the BFLNM, physical, emotional, and mental side effects that a few short years of desperate dieting will provoke can be prevented. I have selected this young age group because they are the formative years upon which lifetime habits are formed, and further because research by the American Dietetic Association indicates that girls as young as nine are currently dieting, with many more joining them in the teenage years (Duyff, 2006, p. 434). In addition, many girls—“about 50 percent of nine to fifteen year olds—see themselves as overweight” (Duyff, 2006, p. 434). Young dancers adopt the attitudes of the older girls with whom they dance. Therefore, intervention must be early.

Contemporary dancer Katharine Vigmstad (2011) wrote the following about pressures she felt to achieve a certain body type:

“Comparison with other dancers is encouraged from the beginning, and this early cultivation of ‘what a dancer looks like’ into the minds of young dancers becomes an integral part of their notion of self-worth and subsequently, their identity. The objectification of a dancer comes from teachers and spectators as well.” (p. 2)

Those outside the dance community may easily say the ideal dancer body—what instructors and directors promote as essential for those serious about the art—is the problem. While this pressure is genuine and affects dancers from an early age, the answer primarily lies not in the problem of dancers desiring a thinner body, but in the lack of proper education when dancers desperately work to obtain that ideal body weight. While many dance instructors work to decrease the amount of emphasis on body weight or shape, the simple fact is that a

dancer's body ideal will not vary much—just as it has been consistent in the past 40+ years. If it is possible to protect a dancer's health, physically, mentally, and emotionally in achieving that body, a preventative measure must be taken. She may be saved a lifetime of heartache and self-destructive behavior. The BFLNM is the way to achieving this goal, giving dancers the knowledge to enable them in changing their eating patterns healthfully to lose weight and body fat percentages.

Many of current ideas in academia propose that instructors and choreographers avoid “encourag[ing] or perpetuat[ing] the body perception distortions with constant admonitions to lose weight” (Culnane and Deutsch, 2011, p. 95). This type of approach, however, while working to decrease external pressures for a dancer to lose weight, is ultimately defeated because it does not address the internal—and much more pronounced—pressures for a dancer to lose weight. Once again, the problem is not primarily in a dancer believing she should lose weight; the problem is that when she does want to lose weight, she does not know the proper way to do so. Such an effort as appropriate chiseling of muscles, diminishing of body fat, and decreasing body weight exists. This is possible through an appropriate, healthful manner, protecting not only the dancer's physical health, but also her psychological and emotional health. “Eating disorders are linked to psychological problems ... the person's whole life—schoolwork or career, family life, overall health—gets wrapped up in the eating issues” (Duyff, 2006).

Because the model has been proven to be an aid for females to reach the low-weight, low-body fat goals they desire, dancers will consistently comply with the lifestyle. Older dancers who have already developed improper eating habits may also benefit from the BFLNM. Although their habits—particularly cognitive cyclic patterns of dieting and self-punishment for any lack of adherence to self-established rules—are already well founded, the education of this model may help them adjust their lifestyles to healthier weight-loss or maintenance programs. Future studies are necessary to indicate whether or not a dancer over the age of 18 may comply with the BFLNM, based on her preconditioned understanding of diet and weight control.

BFLNM Defined

The BFLNM is the “Eating for Life” section taken from the book *Body for Life: 12 Weeks to Mental and Physical Strength*. The book, published in 1999, educates the public on body transformation and features a 12-week competition developed by former competitive body-builder Bill Phillips. While the diet is one on which “healthy adults, age 18 and over” may lose considerable body fat and gain lean muscle, this

is only in combination with the intense exercise program. Power weight-lifting principles are inapplicable to most types of dancers who desire a leaner, wispier appearance. Therefore, I am only advocating the nutrition element of BFLNM. I chose this model because it best packages scientifically researched, successful elements of nutrition and weight loss into an easy-to-follow lifestyle. While BFLNM was not developed for dancers or adolescents under the age of 18, the following supportive evidence will indicate why becoming educated about this nutrition model is best for this demographic of young dancers. Even if they do not apply it at a young age, there will come a time in their late teenage years or even early twenties that the education will prove preventative.

The National Institute of Health (2011) states, "Although there are many programs advertised to help [individuals] lose weight, the only proven long-term and safe method is to burn more calories than you consume." Phillips (1996) states in his book that, "results will differ, even when using the same program" (v). This is dependent on the predisposition of an individual's genetics. Each dancer on the BFLNM will not see the same results but *will* see herself reach her own body's best appearance. Average weight loss based on the featured success stories averages around two to three pounds of weight loss per week. Healthy weight loss is slow because a true caloric deficit must be reached to eliminate a pound of body fat, not water weight. For this reason, a dancer needs to understand that weighing herself every day has a negative psychological effect because she will not see daily results. This program is efficient only upon daily adherence. While Phillips promotes a 12-week program in his book, a dancer may and should adopt this model as a lifestyle because of its well-rounded programming.

The concepts explained in the nutrition model are those that will best enable a dancer to lower her body fat percentage without destroying essential lean muscle and performance endurance. The elements of the BFLNM are simple: caloric deficit; frequent, smaller meals; protein and carbohydrate combinations; vegetable and fruit consumption; taking a multi-vitamin; consistent hydration; and one "free day" per week.

Caloric Deficit

For weight loss to occur, an individual must have a daily caloric deficit. In other words, this means more calories out than in. Low caloric intake promotes fat loss while providing enough calories and protein to maintain muscle and cardiovascular endurance. The American Dietetic Association recommends girls ages 9–13 eat 1,600–2,200 calo-

ries daily, while teenage girls ages 14–18 should eat 1,800–2,400 calories daily (Duyff, 2006). Cutting the younger group's intake to 1,300 will provide anywhere from a 300–1,000 caloric deficit, while cutting the older group's intake to 1,500 will provide anywhere from a 300–900 caloric deficit. Because losing one pound of body fat requires a deficit of 3,500 calories (Duyff, 2006), these numbers equate to a healthy weight loss of one half to two pounds per week without any additional exercise.

Frequent, Smaller Meals

The BFLNM encourages eating five to six small meals throughout the day. Frequent, smaller meals also keep blood sugar, or glucose, at consistent levels. This prevents insulin from turning excess glucose into glycogen that the liver and muscles cannot store. To sum this process up, eating six small meals per day instead of three larger meals prevents the body from storing body fat. Calories should be evenly divided among the smaller meals, which should be eaten “every two to three hours” (Phillips, 1999, p. 98).

Andros (2010) states that those serious about weight loss need to “eat smaller and more frequent meals of quality calories” and “must never miss meals” (p. 257). Missing meals causes blood-sugar levels to fall, say nutritionists Aceto and Velazquez (2007), initiating binge-inducing increase in appetite. “Continual influx of amino acids also blunts appetite, makes you feel fuller longer, which helps control calorie consumption and body fat” (p. 186).

Protein and Carbohydrate Combos

BFLNM states that each small meal should contain a “portion of protein and carbohydrates,” where a portion is defined as “the size of the palm of your hand or your clenched fist (Phillips 1996, p. 98, 133). Examples include steak and rice, tuna and wheat bread, or salmon and a baked potato. This combination stimulates the body's natural fat-burning hormones of glucagon and insulin. Protein produces glucagon while carbohydrates produce insulin, the combination of which fills the bloodstream with traffic that slows down glucose. When glucose moves through the bloodstream too quickly, blood sugar levels will dip, leaving an individual with sugar cravings or deficient of energy (Dworkine, 1998, p. 60).

On a restricted calorie diet, individuals burn not just body fat but also muscle tissue. By eating protein-rich food (in combination with carbohydrates) at each meal and snack in between, dietician Nancy

Clark says that it will help keep an individual “feeling fed as well as curbing [their] appetite (Clark, 1997, p. 38).

While many weight-loss diets focus on a high-protein, low-carbohydrate intake, the American Dietetic Association frowns upon the approach because it offers inadequate levels of glucose for brain function. “The Adequate Intake Level (AI) is based on the lowest amount of glucose (blood sugar) needed daily for normal brain function. That is a minimum of 130 grams of carbohydrate, which equals 520 calories” (Duyff, 2006). Glucose is the only form of energy the brain can use, and carbohydrates are the main source for glucose. However, many popular diets go to the other extreme of eliminating protein, which has the myth of leading a dancer to pack on more muscle. On the contrary, a higher protein intake—nearly equal with carbohydrate intake—is essential when in a caloric deficit. Duyff (2006) of the ADA said, “The fewer calories you consume overall, the greater percent of calories needed from protein to meet your protein need.”

Vegetable and Fruit Consumption

The BFLNM states, “Add a portion of vegetables to at least two meals daily” (Phillips, 1999, p. 98). It also encourages fruit to be added as carbohydrates in several meals per day.

The Harvard School of Public Health (2011) advocates eating fruits and vegetables for their abilities to lower blood pressure, lower risk of heart disease, and provide adequate nutrients for healthy growth and development in teens.

MyPyramid.gov (2011), a website for the United States Department of Agriculture, encourages individuals to eat a variety of fruits and vegetables. Dietary fiber from fruits, as part of an overall healthy diet, helps reduce blood cholesterol levels and may lower risk of heart disease. Fiber is important for proper bowel function. It helps reduce constipation and diverticulosis. Fiber-containing foods such as fruits help provide a feeling of fullness with fewer calories.

Taking a Multi-Vitamin

American Dietetic Association studies indicate that teenagers as well as adults are healthier and get the daily nutrients required for optimum physical performance when taking a multi-vitamin (Duyff, 2006).

Quite logically, the less fuel a dancer ingests the fewer vitamins that she consumes. This can become problematic when a dancer is restricting her caloric intake. Although the BFLNM provides guidelines to ensure essential nutrients are consumed through the diet, consump-

tion of a multi-vitamin once daily may, as the Harvard School of Public Health (2011) states, “replace some of those losses.”

Consistent Hydration

The National Institute of Health (2011) indicates that drinking about 80 ounces of water per day is healthy, whereas the BFLNM suggests drinking one to two cups of water (8–16 ounces) before every meal. This equates to a range around the suggested intake amount. Chris Aceto, a nutritionist for bodybuilders, states that “mildly dehydrated individuals demonstrate up to a 2% decline in metabolism (Aceto and Velazquez, 2007, p. 185). For a dieting dancer, a decline in metabolism means less caloric output, which equals less weight loss. Furthermore, a 2003 study from the *Journal of Clinical Endocrinology & Metabolism* found that drinking half a liter of water raised metabolism by 30% (Brown).

Daoust and Daoust (2001) confirm Aceto and Velazquez’s encouragement of consistent hydration. The couple, who are also nutritionists, also encourage that a dieter “avoid caffeinated beverages and only drink 1% low fat milk can be used as part of your meal or as a snack.” The Dietary Reference Intakes from the National Institute of Health (2011) advise an Adequate Intake (AI) level of 2.7 liters (91 ounces daily). A liter is about 1 quart or 4 cups. Teens and children need somewhat less.

One “Free Day” Per Week

The BFLNM states that a “free day” is “a day on the program when you should eat the foods you’ve craved the previous six days” (Phillips, 1999, p. 131). While the individual is encouraged to eat foods usually termed ‘bad,’ this day should not be a full-out binge day. A caloric top is not used—the dieter must use her own discretion. The rationale in doing so is supported by a study conducted by White et al. (2009), who found that “induction of a calorie-reduced diet leads to decreases—rather than increases—in binge eating” (p. 326).

Furthermore, a dancer’s body will quickly adapt to the number of calories she puts in it. When running a caloric deficit, she will only lose weight for so long before her metabolism slows to this level. The best way to avoid this leveling off is to include a free day. The free day resets the metabolism so the body does not get used to the same caloric load. The reason for the free day is consequently both physical and psychological—it allows a dancer to overcome cravings and reset her metabolism so she gets leaner during the week.

A clinical study published in the *Journal of Sport Nutrition and Exercise Metabolism* in 2006 (Candow) identified the BFL program's success. The study compared three groups of adult men and women who were overweight or obese. The first group followed the BFL plan, the second followed the American Heart Association's recommendations, and the control group followed a typical American sedentary lifestyle. Results after 12 weeks revealed that the BFL group lost approximately double the amount of body and stomach fat and had lower bad cholesterol levels and lower blood pressure values in comparison to the other two groups.

Instructors' and Educators' Role

For instructors and educators to be effective in their role of presenting information to their dancers, it is imperative to first become educated themselves. Unless one who has the potential to influence hundreds of dancers becomes excited about the possibilities for protecting the future health of their dancers, they will be unwilling to follow the suggested educational procedures.

Oliver (2008) advocated that dance educators have a primary responsibility to help their students in these avenues. She said, "By listening and responding to student concerns, and by being open to changing our pedagogical methods, we can create an environment in which a positive body image is more readily attainable" (p. 25).

Instructors and educators can do just that by learning about the BFL model through several avenues: through purchase of the BFLNM book, which is available at book retailers everywhere in addition to public libraries; through the BFLNM website at www.bodyforlife.com; or through a website that I will build for the very purpose of educating instructors and educators on how to educate their dancers. This website will include a downloadable PowerPoint presentation for them to utilize in their parent/child workshops that will be outlined below. In addition, the website will display numerous links to articles that address components of nutrition that will be of further aid to them in educating their dancers, whether through weekly handouts or class-time discussions. Resources are widely available. The International Association for Dance Medicine and Science holds conferences for dancers and educators on topics including nutrition. The National Dance Alliance holds conferences and features publications on dance nutrition and injuries. In addition, popular dance magazine like *Dance Teacher* and *Dance Magazine* regularly feature articles on nutrition and injury prevention. Resources for instructors and educators to use in teaching their adolescent dancers are abundant.

How to Educate Young Dancers

In order for the BFLNM to succeed in preventing eating disorders, instructors (private studio/company setting) and educators (public school setting) need to commit to several ideas: a biannual parent/child nutrition workshop; hand out weekly articles on nutrition to their dancers; and reserve five minutes at the beginning of class to discuss the nutrition topic of the week.

Convenient times for biannual parent/child nutrition workshops are in August and January, the beginning of fall and spring terms. Because most dance studios and schools in the U.S. focus on the traditional school-year calendar, these months are when new semester classes begin. Nutrition workshops should be mandatory for each dancer and a parent or guardian to attend to participate in the semester's classes. At the workshop, the educator or instructor should present the basics of healthful eating for dancers, based on the BFLNM.

As instructors and educators hand out weekly nutrition "readings about body image and nutrition" that encourage proper and healthful eating habits and "discuss them in class," dancers will continue their health education throughout the course of the semester (Oliver, 2008, p. 24). In an academic setting, these readings may be mandatory with a follow-up quiz to ensure students read and understand the articles. In lieu of a quiz, Oliver (2008) suggests that "students can also write short, informal papers commenting on their perceptions of their own body and of themselves as dancers" (p. 25). However, in a private dance studio setting, mandatory reading for dancers may be an ineffective tool because of lack of an enforceable policy. Instead, instructors could simply encourage their students to read the weekly article. These articles could additionally be e-mailed to dancers' parents to keep the entire dancer's family aware of proper nutrition and health.

While dancers will ideally read each week's nutrition article before attending class the following week, reserving five minutes at the beginning of class to discuss the nutrition topic of the week will ensure each dancer digests the information, whether she read it or not. If she read the article, the discussion will further enforce the principle of focus, while if she did not read the article, she will have a time to hear the information, and potentially receive positive feedback from other dancers who did, in fact, adhere to the reading schedule.

An Element to Consider

While critics may argue that by exposing young minds to such things as the rigors of nutrition, caloric balancing, and meal timing, young dancers will become preoccupied with their eating before they

may ever feel naturally inclined, with Griner's (2006) statistic of 50% of dancers found to have been engaging in eating disorders at some point in their dancing years, the statistic declares that more positive prevention will result than negative influence (Gordon, 1981, p. 45). In no way should the advocated nutrition workshops be focused on eating disorders or any stigmas of which these young dancers may already be aware. Instead, workshops should revolve around the outcomes of BFLNM—adequate nutrition for optimum performance, maintaining or reaching healthy, yet lower weights, immune system protection, and injury prevention.

While these workshops may be helpful for dancers already struggling with disordered eating, they are primarily a preventative procedure. As Griner (2006) states, "Prevention means providing education and information to increase understanding, enhance attitudes, and promote healthier behavior" (p. 45).

BFL Training Will Answer Dancers' Questions

As a nationally certified personal trainer in addition to my own dance training and performance, the numbers of dancers that approach me for nutrition advice and weight loss are staggering. Typical questions dancers ask include the following: "How can I drop five pounds? I need to look good for my concert," "How can I get rid of this fat right here? I hate it and it makes me so mad to see it in the mirror," and "What can I eat to give me more energy in class? I feel like I'm dragging." These three very frequently asked questions are answered by the BFLNM.

How can I drop five pounds?

For a dancer to lose five pounds, she must have a caloric deficit of 17,500 calories (3,500 calories equates to one pound). While exercise accelerates deficit levels, reduction of caloric intake through dieting is fundamental. The BFLNM nutrition model focuses on a 1,500-calorie daily diet, implementing small, frequent meals spread throughout the day that revs calorie-burning metabolism and leads to greater weight loss because of the minimal amount of food in the stomach at one time.

How can I get rid of this fat right here?

No such effort like spot reduction exists when attempting to change one's physical appearance (Hamilton, 1995, p. 49). Genetics determine areas where an individual's body holds its fat. While exer-

cise may tone the underlying muscle, only balanced nutrition featuring caloric deficit will trim body fat percentages. However, each person's body draws upon its genetic blueprint to determine what part of the body from which adipose cells (fat) are reduced. This is a potential breeding ground for dancers to resort to self-starvation to force the body to reduce its body fat. The BFLNM nutrition model provides a nutritionally sound diet, while still working to decrease the body's storage of fat to minimal levels. This model will also slim those trouble areas for good—not temporarily as long as a dancer maintains the lifestyle. Other unhealthy methods may temporarily remove the unwanted fat, but allow it to come back.

What can I eat to give me more energy in class?

Energy is commonly known to come in the form of calories; however, if a dancer is eating a restricted number of calories, how can she maintain her energy levels? The BFLNM nutrition model functions on amino acid theories—that to provide the adequate amount of amino acids that fuel the body with energy, a high-quality protein source must be combined with a high-quality carbohydrate source. This satiates the appetite and allows the body to reach its highest possible energy levels for physical performance and athletic endeavors.

Twenty-year-old Lesley Pinder is a prime example of how the BFLNM works. Although not a dancer, 12 weeks of following the BFLNM nutrition model allowed Lesley to reduce her body fat percentage from 23 to 13%. Another example is 29-year-old Jen Weatherman, who dropped her body fat percentage from 25 to 12%. The BFLNM website boasts scores of these transformation stories. This program dropped these women's body fat percentages down into what the American Dietetic Association states as the range of essential body fat for women (8–13%) to maintain bone functionality and more critically, avoiding amenorrhea—loss of the menstrual cycle (Duyff, 2006). In Griner's (2006) report of a study in dancers' body fat percentages; she stated professional modern dancers had an average body composition of 12.2% body fat, where the average body composition of female dancers at in universities from different styles of dance was found with an average body composition of 14.2–14.7% (Calabrese & Kirkendall, 1983, p. 51).

Conclusion

What dancers and non-dancers view as the ideal dancer image will not change. It has not changed over the course of the past four dec-

ades, and evidence of a future change is nonexistent because of the media's exaggerated emphasis of beauty as thinness. Because of this increased emphasis, dancers as young as nine will be affected by both internal and external pressure to attain that body shape. With high statistics of perfectionism and achievement standards, this puts young dancers at risk for developing disordered eating patterns that may afflict them physically, emotionally, and mentally for the rest of their lives. However, when educators and instructors of young dancers in their formative years, ages nine through eighteen, implement the education on the BFLNM nutrition model, these dancers will have a thorough understanding of properly and healthfully attaining that body shape. Genetics and metabolism play a major role in determining a dancer's body shape and size; however, this does not mean a dancer needs to accept her body that is at variance with the ideal. She *can* attain her genetic predisposition's ideal body, and *will* with proper education on the BFLNM nutrition model—a safe, effective, and proven nutrition model. BFLNM is not a diet—it is a lifestyle.

Dance instructors need to lead this effort—an effort to educate dancers about the proper, healthful way to achieve those goals. This may be the greatest step in impacting future generations of dancers' health. Proper knowledge of eating for prime aesthetic appearance will protect a dancer's physical, psychological, and emotional health for her entire lifetime. By learning about the BFLNM nutrition model from their dance instructors and educators, dancers will engage in healthier, more nutritious lifestyles while reaching their goals in aesthetic appearance.

Works Cited

Aceto, Chris, and Velazquez, Eric. 2007. Never diet again. *Joe Weider's Muscle & Fitness*, 68, 184-192.

Andros, Sean. 2010. Diet dangers. *Flex*, 254-264.

Braisted, Jacquelyn R.; Mellin, Laurel; Gong, Elizabeth J.; Irwin, Charles E. 1985. The Adolescent Ballet Dancer: Nutritional Practices and Characteristics Associated with Anorexia Nervosa. *Journal of Adolescent Health Care*, 6:5, 365-371.

Brown, Clive M.; Dulloo, Abdul G.; Montani, Jean-Pierre. 2003. Water-Induced Thermogenesis Reconsidered: The Effects of Osmolality and Water Temperature on Energy Expenditure after

Drinking. *Journal of Clinical Endocrinology and Metabolism*. 91: 3598-3602

Brooks-Gunn, J. J., C. Burrow, M.P. Warren. 1988. Attitudes toward eating and body weight in different groups of female adolescent athletes. *International Journal of Eating Disorders*. 7, 749-757.

Candow, Darren G.; Burke, Natalie C.; Smith-Palmer, T.; Burke. 2006. Effect of Whey and Soy Protein Supplementation Combined With Resistance Training in Young Adults. *International Journal of Sport Nutrition and Exercise Metabolism*. 16:3.

Clark, Nancy. 1997. Protein: how much. *Dance Teacher Now*. 38-40.

Culnane, Catherine, and Deutsch, Donna. 2011. Dancer disordered eating: comparison of disordered eating behavior and nutritional status among female dancers. *Journal of Dance Medicine & Science*. 2, 94-100

Daoust, Gene and Joyce. 2001. *The formula: a personalized 40-30-30 weight loss program*. New York: Ballantine.

Duyff, Roberta Larson. 2006. *American Dietetic Association Complete Food and Nutrition Guide*. Hoboken, New Jersey: John Wiley & Sons.

Dworkine, Norine. 1998. Eat to win. *Vegetarian Times*. 245, 60.

Gordon, Suzanne. 1983. *Off Balance: The Real World of Ballet*. New York: Pantheon.

Griner, Brenda, et al. 2006. Achieving the perfect body: nutritional behaviors of nonprofessional, regional female dancers. *Journal of Physical Education, Recreation & Dance*. 77, 40-45, 51.

Hamilton, Linda. 1995. The dancer's dilemma: how to be healthy and thin. *Dance Magazine*. 48-49.

Harvard School of Public Health. 2011. <http://www.hsph.harvard.edu/nutritionsource/what-should-you-eat/vitamins/>

Kirkland, Gelsey. *Dancing on my Grave*. 1987. Jove Books.

Montanari, A, Zietkiewicz, EA, Montanari. 2000. "Adolescent South

African Ballet Dancers" South African Journal of Psychology. 302, 31-35,46.

National Institute of Health. 2011. http://www.nlm.nih.gov/medlineplus/news/fullstory_121882.html

Oliver, Wendy. 2008. Body image in the dance class. *Journal of Physical Education, Recreation & Dance*, 79, 18-25.

Phillips, Bill. 1999. Body for life: 12 weeks to mental and physical strength. New York: HarperCollins Publishers.

Phillips, Bill. 1996. Body for Life. <http://www.bodyforlife.com>

Ravaldi, Claudia; Vannacci, Alfredo; Bolognesi, Enrica; Mancini Stefania; Faravelli, Carlo; Ricca, Valdo. 2006. Gender role, eating disorder symptoms, and body image concern in ballet dancers. *Journal of Dance Medicine and Science*, 61, 529-35.

Schluger, Alice E. 2010. Disordered eating attitudes and behaviors in female college dance students: comparison of modern dance and ballet dance majors. *North American Journal of Psychology*, 12,117-128.

United States Department of Agriculture. 2011. http://www.mypyramid.gov/pyramid/fruits_why.html

Vigmostad, Katharine. 2011. Dancing through boundaries. *The Axis Syllabus*.

White, Marney A., et. al. 2009. Regimented and lifestyle restraint in binge eating disorder. *International Journal of Eating Disorders*. 42, 326-331.

Yannakoulia, M., Sitara, M., & Matalas, A. (2002). Reported Eating Behavior and Attitudes Improvement after a Nutrition Intervention Program in a Group of Young Female Dancers. *International Journal of Sport Nutrition & Exercise Metabolism*. 12(1), 24-32.

Motivational Behaviors of Elite Senior Athletes

M. Vinson Miner and Bret H. Boyer
Utah Valley University

Abstract

Qualitative research methods were utilized to analyze, compare, and develop insight into just how extrinsic motivation can affect intrinsic motivation. Extrinsic motivation moves from being more external in nature to a point where motivation is identified, integrated, or aligned with intrinsic motivation. An external factor could affect and help explain why elite senior athletes are motivated to engage in, participate, and compete in senior games and competitions. A ten-point survey/questionnaire was designed to illuminate the unique characteristics and commonalities associated with the participants. Qualitative components were observed, analyzed, and documented. (External rewards enhance motivation 77.5% responded in the affirmative). A variety of questioning techniques were employed including a Likert scale and fill-in-the-blank/open-ended questions. Participants' psychological responses, preferences, likes/dislikes, and attitude were tabulated and compared. In conclusion, the majority of respondents indicated that extrinsic rewards (on a scale of 1 to 10, average score was 7.58) were clearly important elements in their internalized or intrinsic motivation to engage, participate, and compete in Elite Senior World Games.

Introduction

Research Question: To what degree does extrinsic motivation (the receiving of an award, medal, trophy, praise, or social approval) affect intrinsic motivation (internal self-determination, or being motivated from within, to be self-motivated) to compete in the World Senior Games.

As part of an on-going investigation into the elusive events behind elite senior athletes' motivation and an ever-increasing number of elite senior athletes pursuing and participating in organized competitive physical activities; additional research is needed and would help illuminate some important aspects of the just how extrinsic motivation (like the receiving of an award, medal, trophy, praise, or social approval) affects their intrinsic motivation (self-determine or self-motivated) to compete and achieve their competitive goals.

The purpose of this study was to examine the dynamic relationship that may exist between various facets of extrinsic motivation (the receiving of an award or medal) that affects elite senior athletes' intrinsic motivation (self-determination or self-motivation) to compete in organized senior games. Within those relationships, there are unique connections that could positively transfer motivational power that is deeply ingrained in each elite senior athlete.

Methods

Design of the Study

Data for this case study to develop grounded theory were collected over a 2-week period at the World Senior Games in St. George, Utah, held October 3-16, 2010. The research sample consisted of a ten-point survey/questionnaire designed to illuminate the unique characteristics or commonalities associated with the participants. (n=116). Questions consisted of yes or no answers, degree or Likert-scale type, and fill-in-the-blank/open-ended questions. The questionnaires were personally administered by the researchers.

Results

Question 1: On a scale of 1-10 (1 being least motivated, 10 being most motivated), does receiving an external reward, like a medal, positively impact your intrinsic motivation?

The average answer (score) was 7.58.

Question 2: On a scale of 1-10, how important to you is receiving a reward or medal have on your desire to participate?

The average answer (score) was 6.95.

Question 3: On a scale of 1-10, would you be proud of your medal, wear it around, or even who it off as a positive way to enhance your continual motivation or participation?

The average answer (score) was 5.90.

Question 4: In your opinion, can an external reward like a medal or trophy actually enhance your motivation to participate?

90 responded Yes, and 26 responded No.

Question 5: Would you participate in events like this if you didn't have the opportunity to receive a reward or medal?

103 responded Yes and 11 responded No.

Question 6: Does the opportunity to win a reward or medal enhance your self-determination and self-desire to participate?

88 responded Yes and 27 responded No.

Question 7: Does winning a medal build your self-confidence and your desire to participate again in such events?

82 responded Yes and 34 responded No.

Question 8: Do you see the winning of medal as an incentive to participate?

78 responded Yes and 37 responded No.

Question 9: How many medals have you won in elite senior Athletic events?

Bronze: 190

Silver: 173

Gold: 226

Total medals: 589 (10 outliers were not included in the final tabulation, thus divided by 106 participants=average per participant=5.5).

Question 10: Does receiving a medal for winning add or detract from participations in athletic events?

114 responded that it adds and 2 responded that it detracts from motivation.

Discussion

This research provided information and insight and helped to develop grounded theory into why senior elite athletes participate in phys-

ical activities and competitions. It may help institutions, organizations, and other senior groups to design their rewards programs more efficiently to meet the specific motivational aspects of elite senior athletes. This valuable information could in turn help motivate more elite senior athletes to enter, compete, and participate in athletic contests such as the World Senior Games.

It also serves to promote an increased awareness with each senior elite athlete about their unique motivational characteristics as it applied to seeking a healthy lifestyle, thus enhancing their intrinsic motives and helping them tune in key aspects that could help them achieve their degrees. In addition, the information and insights gained could help instructors, coaches, athletes, and seniors understand the unique intrinsic motivational aspects of each senior elite athlete and incorporate plans to motivationally assist them in their quest to achieve their fitness activity and competition goals.

Conclusion

This study clearly demonstrates that extrinsic rewards, like medals, trophies, praise, or social approval, can contribute and affect in a significant way intrinsic motivation. The study also illustrated common themes, patterns, and categories as highlighted by the respondents' answers to the questionnaire. There is a dynamic relationship and connection between the receiving or opportunity to receive an extrinsic reward and the internal or self-motivation of elite senior athletes to participate in senior games.

References

- Deci, E.L., Koestner, R., and Ryan, R. M. (1990). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivations. *Psychological Bulletin*, 125, 627-668.
- Deci, E. L., and Ryan, R. M. (1985). *Intrinsic Motivation and self-determination in human behavior*. New York: Plenum.
- Deci, E. L., and Ryan, R. M. (1991). A motivational approach to self: integration in personality. In R. A. Dienstbier (Ed.) Nebraska Symposium on Motivation 1991: Vol 38. *Perspectives on motivation: Current theory and research in motivation* (pp. 237-288) Lincoln, NE University of Nebraska Press.

Deci, E.L., Vallerant, R. J., Pelletier, L.G., and Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *The Educational Psychologist*, 26, 325-346

Fortier, M.S., Vallerand, R.J., Briere, N, M., and Provencher, P.J. (1995). Competitive and recreational sport structures and gender: A test of their relationship with sport motivation. *International Journal of Sport Psychology*, 26, 24-39

Frederick, C. M., and Ryan, R. M. (1995). Self-determination in sport: A review using cognitive evaluation theory. *International Journal of Sport Psychology*, 26, 5-23

Lepper, M. R., and Greene, D. (1975). Turning play into work: effects of adult surveillance and extrinsic rewards on children's intrinsic motivation. *Journal of Personality and Social Psychology*, 31, 479-486

Pelletier, L. G., Fortier, M. S., Vallerand, R. J., Tuson, K. M., Briere, N. M., and Blais, M. R. (1995). Toward a new measure of intrinsic motivation, extrinsic motivation, and amotivation in sports: The sport motivation scale (SMS). *Journal of Sport and Exercise Psychology*, 17, 35-53

Vallerand, R. J., and Losier, G. F. (1999). An integrative analysis of intrinsic and extrinsic motivation in sport. *Journal of Applied Sport Psychology*, 11, 142-169.

Walling, M. D., Duda, J. L., and Chi, L. (1993). The perceived motivational climate in sport questionnaire: Construct and predictive validity. *Journal of Sport and Exercise Psychology*, 15, 172-183.

Oedipus's Family in José Luis Borau's *Furtivos* and Icíar Bollaín's *Flores de otro mundo*

Matías Martínez Abeijón
Southern Utah University

Abstract

This article aims to compare the manner in which two films, Furtivos (1975) directed by José Luis Borau and Flores de otro mundo (1998) by Icíar Bollaín, focus on the representation of a non-traditional family living in the Spanish countryside. Using as a departure point the metaphorical readings prevalent in the critical approach to Furtivos, I explore the meaning of the similarities between both plots in their representation of the family (i.e., the presence of a strong Oedipal tension that will come to be disturbed by the arrival of an outsider, the family as a microcosm for larger social and national issues) in the historical contexts in which these films were produced. I also analyze the historical significance of revisiting during the democracy the metaphors that had been previously used during General Francisco Franco's dictatorship. From a theoretical point of view, this discussion studies the presence of those clichés traditionally associated with film genre (i.e., melodrama's emphasis on the family table) while making reference to their relationship with the notions of gender and power.

When it comes to the critical analysis of *Furtivos* [*Poachers*] (1975) by Spanish director José Luis Borau, it becomes difficult to separate the realms of history, politics, and the symbolic. After a quarrel between the censors and the production company, the film had an eventful release in the San Sebastián Film Festival, less than three months before the death of General Francisco Franco. Spanish film critics Esteve Riambau and Carlos Fernández Heredero called this time “the years of the metaphor” (Riambau “cine” 109), a historical moment in which “the fight against censorship polarizes a great deal of the energies dedicated to create film in a democratic and politically conscious manner” (Fernández Heredero 741).¹ As a consequence, the study of *Furtivos* has been shaped by analogies between the storyline of the film and the politics of the time. Thus, the Oedipal relationship between the main characters, Ángel, a poacher, and Martina, his mother, has been interpreted as a straightforward metaphor for the relationship between the Spanish population and the Franco regime. At the time, family was frequently presented as a microcosm of the nation’s own situation. This kind of reading was encouraged by the director himself, who told the critics that he had also drawn inspiration from a female version of the mythological figure of Saturn devouring his children, created by Spanish filmmaker Luis Buñuel (Fernández Heredero 740; Kinder 234). Furthermore, the presence in the plot of such an authority figure as the loosely named “governor,” with his privileged access to restricted hunting grounds, opened the door to a direct association between what the spectators were watching on the screen and the inequalities that characterized the Franco regime.²

A quarter of a century later, the film *Flores de otro mundo* [*Flowers from Another World*] (1999) by Spanish director Icíar Bollain followed three unlikely couples in an inhospitable rural environment not so different from the setting previously used for the action of *Furtivos*. The storyline is vaguely related to an actual event that occurred in a small town in the remote mountains of Huesca. There, a group of men

¹ [... los años de la metáfora] (Riambau 109); [... la lucha contra la censura polariza buena parte de las energías del cine consciente y democrático que se produce en España] (Fernández Heredero 741).

² Manuel Gutiérrez Aragón, one of the screenwriters for the movie, has discussed the importance of this character for understanding the political environment in which the movie was released. “A few months earlier, we couldn’t have included someone like the governor, and a few months later, after the death of Franco, it wouldn’t have mattered any more whether he had been in it or not [...] People saw the governor as representing Franco’s power, the Frankist political class” (Besas 135-136).

were in a local bar watching *Westward the Women* by William Wellman (1951), when they decided to organize a similar trip to their own town (Marín). In Bollain's movie, one of the couples resulting from the arrival of the women is the one formed by Damián, a local, and Patricia, an outsider. Mirroring the plot of *Furtivos*, Damián is a single man who has difficulties communicating and lives with an authoritarian mother, Gregoria, in an environment characterized by the modesty of their means. Regardless of which one of the two movies we are discussing, poverty is evident in the furniture and fixtures of the house. This functions as a commentary on the consumerism embraced by Spanish society from the sixties to the present. Whether we are talking about Martina's home in the seventies or Gregoria's home in the nineties, the two share the same cold nights because of the lack of heat and the same drab clothing that lacks any sense of fashion or beauty. Furthermore, the dialogue in *Flores de otro mundo* reveals to the audience that even the most basic shopping at the supermarket may be considered an inappropriate excursion into the realm of luxury and, thus, may become susceptible to harsh criticism from the matriarch. Keeping a sense of decorum with this poverty, the film renounces any kind of technical sophistication that may call attention to itself. The audience is left alone with uneventful camera work, which underscores the realistic sobriety of the universe that the movie chooses to portray.

The similarities between the films also include the irruption of a second and younger female character: Milagros in *Furtivos* and Patricia in *Flores de otro mundo*. Their arrivals represent a challenge for the status quo of the relationship between mother and son. Consideration of the two films in relation to each other in this way allows both Bollain and the audience to verify that the processes of modernization and democratization have not necessarily found a solution for some of the traditional issues of living in the country. On another level, it emphasizes the manner in which Bollain intends to humanize the debate about immigration, which was starting to become a serious issue of public debate in Spain in the late 1990s. Patricia, played by Dominican actress Lissette Mejía, is presented as an outsider with a problematic past, which is exactly the same manner in which Borau characterized Milagros in *Furtivos*. Patricia's condition as a foreigner and an immigrant is hence made secondary to her status as an outsider.

Discussing at any length the value of the Oedipal myth for psychoanalysis or post-structuralist theory is beyond the scope of this in-

quiry.³ It is clear, however, that the myth of Oedipus and his mother, Jocasta, has frequently served as a catalyst for the discussion of social, psychological, and political issues because of what film scholar Marsha Kinder characterizes as its “potential to function subversively as a vehicle for cultural change” (Kinder 197). It seems that this kind of debate was what Buñuel and Borau, among others, were trying to elicit when they chose such a subject in the 1970s. In bringing back the same topic to the threshold of the twenty-first century, Bollaín injects new life into the myth while carrying over some of the associated clichés and mannerisms. When analyzing its presence in the history of film in Spain, Kinder reminds us of the manner in which the Oedipal narrative, “through its successful proliferation and compulsive repetition, [...] helps the dominant patriarchal culture reproduce itself”—a reading that, following the release of *Furtivos* in the middle 1970s, was espoused by then prevailing post-structuralist critical writing. In general, the incestuous nature of the relationship between the mother and the son is usually presented as an anomaly, which, as such, is susceptible to correction. It cannot go unnoticed that this correction rarely is gender-neutral. For example, it does not matter whether our classical source is Homer, Sophocles, or Euripides, the anomalous situation is invariably corrected by the suppression of the mother—that is, by her death, a punishment that is not matched by Oedipus’s fate, which, depending on the tradition, ranges anywhere from exile to the maintaining of his kingdom. Although some of this critical debate may seem anachronistic now, this was an influential view at the time of *Furtivos*’s release. For example, three years earlier, French philosopher and social scientist René Girard had suggested that there was a pattern in myth in which women ended up becoming the scapegoat for all social ills, because of “her weakness and relatively marginal social status. For this reason, she

³ All theory discussed in this article aims to illuminate its subject, that is, the relevance of studying the analogies between these two movies as well as what we can learn from juxtaposing them. There are many areas of inquiry associated with *Furtivos* that have already been thoroughly investigated. When it comes to tracing the historical and literary sources for the film, both Agustín Sánchez Vidal and Peter Evans are invaluable sources. There is also an abundance of theoretical perspectives, particularly among foreign critics of the film, with the most common theme being the confluence of national identity, gender issues, psychoanalysis, and fairy tales. Evans, Marsha Kinder, and Robin Fiddian discuss in depth the relevance of the critical proposals of such critics as Judith Butler, Barbara Creed, Estela Welldon, Nancy Chodorow, Katya Silvermann, Laura Mulvey, and Adrienne Rich, among others, for the analysis of Borau’s film.

can be viewed as a quasi-sacred figure, both desired and disdained, alternately elevated and abused" (142). How do these connotations translate a quarter of a century later, when Bollaín directs *Flores de otro mundo*? In the meantime, women had become part of both the workforce and the cultural life of Spain. Although a superficial look at the outcome of the Oedipal tension in both triangles might be presented as consistent with such victimization as Kinder and Girard describe, the female characters in both films are constructed with a complexity that transcends such simplification. In both films, women are identities in transition, dynamic characters, capable of terrible injustice and violence, but also victims of their own madness and desire. They are willing to submit to great sacrifices, but seem perpetually forced to live and act in an environment that does not facilitate their agency.

From the point of view of the Oedipal myth, there seems to be a displacement of some traits traditionally associated with the father. The selfishness and tyranny associated with Laius seem actually to characterize both mothers. As Kinder indicates, in those films produced during the dictatorship it is common that "both the source and the target of the violence are literally displaced from the father to the mother" (234). The arrival of the outsider is used to disturb the established balance of power, implicitly presented in both films as anomalous and, as such, susceptible to correction. In *Furtivos*, Martina's inability to accept her new and diminished role ends up making her the victim of Oedipus's rejection and, eventually, also the victim of his violence. Martina "is," at the same time, Yocasta (because of her exposed incestuous relationship with her son) and also Laius (who dies at the hands of his own son). At the end of the movie, the anomaly has been corrected without any other punishment for Oedipus than his loneliness.

In the screenplay by Bollaín and Julio Llamazares, the need to correct this Oedipal tension is emphasized by the repeated pleas made by Damián, asking his mother to accept that Patricia belongs in the house. Gregoria only caves in when Damián himself threatens to leave, which would be an alternative way to correct the anomaly.

Even if violence is not at the heart of all relationships in *Flores de otro mundo*, the balance of power is analogous to that of *Furtivos*. It does not matter whether Gregoria or Patricia ends up reigning in the house, the beneficiary of their activity is Damián—Oedipus. The process always involves the correction of the Oedipal tension, which is characterized as undesirable. When authority is taken away from the older women characters, however, power is not transferred to the younger female outsiders who had seemed to be their antagonists; instead, it goes either to Ángel or to Damián, who act as the referees of the action in both films.

The countryside where the action takes place involves sharp criticism of both the clichés of the rural in the Francoist discourse and the alleged modernization upon the arrival of democracy. In *Furtivos*, the environment in which the action occurs is not what such historians as Sebastian Balfour characterize as Franco's dream of "a peaceful forest" (Evans 117), part of "a mythologized Castilian countryside" (Balfour 266). To the contrary, it is a private hunting ground for those who are entitled to unfair privileges. From the early versions of the screenplay, the governor appears making comical references to how peaceful the environment is (Borau and Gutiérrez Aragón 39) and also how pure the air is (94).⁴ As a contrast, the final version of the screenplay has the local guards concerned because Ángel has apparently killed what they think of as "the Governor's deer" (242).⁵ In *Flores de otro mundo*, the countryside is not a source of good health and pure air either.⁶ Quite the opposite, it enslaves its inhabitants, who seem unable to detach themselves from the land. It does not matter how modern a country Spain becomes, the relationship between the locals and the land seems to derive from the feudal notion that they belong together.⁷ Clearly, in *Flores de otro mundo* violence seems to have been marginalized from the Oedipal setting, yet the same family model remains in place. However, the action also shows the entrenchment of the traditional patriarchal family not only as a cultural institution but also as a pattern of representation.

Bollaín handles this pattern and its anachronistic balance of power by taking advantage of traditional genre motifs of the family melodrama. As a consequence, the movie successfully establishes that cer-

⁴ Evans goes as far as characterizing the movie as a "dystopian pastoral" (117).

⁵ Esteve Rimbau indicates that the fact that the movie discussed the "last" deer points out "the mutual erosion [...] between the rural and the city" [la erosión mutua que se produce entre el mundo rural y la ciudad] ("Escondidas" 133).

⁶ Julio Llamazares, one of the screenwriters of *Flores de otro mundo*, indicates that there seems to be many misunderstandings about what the meaning of living in a rural environment in contemporary Spain (implying that such misunderstandings were one of his motivations for writing the screenplay for the movie) (Bollaín and Llamazares 27). "... the most common view of the town is that of the tourist, the ones that say «how beautiful can that cow be!». But what happens when the Sunday tourists are gone...?" [la visión del pueblo más común es la de turista, la de ¡Qué bonita es la vaca! Pero, ¿qué pasa cuando se los turistas domingueros se van...?] (31).

⁷ On this matter, the scene at the beginning of *Flores de otro mundo* in which Damián indicates that he cannot travel to Cuba because he needs to stay for the harvest season is particularly telling (Bollaín and Llamazares 84).

tain cinematic incarnations of the family are better than others, which is particularly significant when constructing the family as a microcosm for the country, a standard that has been generally accepted by critical readings of *Furtivos*.

Actually, domestic life is at the heart of the competition between the maternal figure and the outsiders in the two films. In *Furtivos*, Martina decides to get rid of Milagros when she starts taking over some of the domestic chores. The young rival dies while she is picking up some potatoes to fix for dinner. In *Flores de otro mundo*, we also see an episode of domestic rivalry involving potatoes, which happens when both Gregoria and Patricia cook and compete for a positive reaction from their companions at the table. Around the dining room table, such an auspicious location for the family melodrama, the masculine character and Patricia's children are forced to act as judges of the women's skills in the kitchen. The mother seems to win this competition when Patricia's children prefer her food. Therefore, Bollain's movie presents an environment in which issues and tensions in the Oedipal narrative are not solved through violence, which is a substantial deviation from the traditions of the myth.⁸ However, this does not show any progress for the place of women in society, as there is not an alternative space to the kitchen for which both women may compete. Just as in many other contemporary revivals of realism, Bollain's film does not seem to grasp the implications of its own reliance on elements of family melodrama. The contradictions entailed by such a choice had been pointed out long ago by Spanish film scholar Román Gubern:

Melodrama is usually [...] deeply reactionary and puritanical. It shows the triumph of virtue and the punishment of vice; it usually reveals the patriarchal structure of power, [it also presents] a rigid and insurmountable separation between social classes (from the point of view of the victim), [it shows] the condition of women as objects and [it showcases] a view of sexuality as sinful.⁹ (288)

⁸ Violence is still associated with gender inequality in *Flores de otro mundo*, albeit in a different context. One of the female outsiders, Milady, has a utilitarian relationship with Carmelo, a local, in which sexual favors work as a commodity until the relationship falls apart and violence erupts. In keeping the violence out of the Oedipal setting, however, it is deprived of both the symbolic dimension and the historical perspective that it has in *Furtivos*.

⁹ [El melodrama suele ser, como queda dicho, profundamente reaccionario y puritano: muestra el triunfo de la virtud y el castigo del vicio, suele reflejar la

A good example of this kind of contradiction can be noticed in the tension created around the axis formed by the dining room table and the bedroom. Jean Baudrillard indicates that, once the characters enter these spaces, they “have as little autonomy in this space as the various family members enjoy in society” (16). The audience also becomes captive, making it easy for them to wish for certain outcomes—such as that Damián and Patricia stop their fighting and stay together. Both films hit a wall, however, when it comes to even hinting at a way in which family relationships might become more egalitarian. Neither film focuses on authority irremediably shifting to Oedipus’s hands. On the contrary, they choose to emphasize how Martina and, to a lesser extent, Gregoria, embody tyranny—a situation that might improve if they were to concede and accept their defeat. When Martina does not, violence arises and she eventually dies. Martina never accepts the correction of the Oedipal tension. From the first version of the screenplay for *Furtivos*, Martina’s character celebrates the disappearance of Milagros, indicating to Ángel that they are better off by themselves (Borau and Gutiérrez Aragón 61). An additional sexual reference to Ángel’s masculinity is added in the second version of the screenplay (Borau and Gutiérrez Aragón 109), adding depth to the Oedipal characterization. On the other hand, in *Flores de otro mundo*, dialogue allows characters to replace the theme of Oedipal violence with the old cliché of “the ages of man” or, more appropriately in this case, “the ages of women.” This shift comes to fruition in a scene set in the cemetery where Patricia is helping Gregoria clean her late husband’s grave. Almost a cliché in recent film produced in Spain, the cemetery scene always allows for raw feelings and provides some kind of historical background for the characters. When Patricia asks Gregoria about her husband, her answer is stripped of any hint of sentimentalism. She says “he was a good man,” which is followed by a rhetorical question by Patricia, who wonders out loud for the audience whether Damián also falls in the same category. In making Damián a younger version of his father, the Oedipal anomaly is corrected. Gregoria becomes a happy grandmother and Patricia, in turn, a younger Gregoria. On the surface, this tacit alliance between both women in *Flores de otro mundo* seems to mirror the idea of the existence of an alternative feminine “genealogy,” as classically characterized by Luce Irigaray and some other thinkers in the orbit of feminism. On the one side, as Latin American anthropologist Marcela

estructura patriarcal del poder, una rígida e infranqueable separación entre las clases sociales (desde el punto de vista de la víctima), la condición de la mujer como objeto y una visión de la sexualidad como pecado] (Mensajes 288).

Lagarde indicates, it seems clear that at this point in the film, "sorority emerges as an alternative to the politics that prevent women from a positive gender identification, recognizing [each other], [and] aggregating [their efforts in an] attuned alliance" ("Pacto" 125).¹⁰ And yet, the effort falls short of even attempting to eliminate the unevenness of the balance of power. Whether we look at Gregoria's past or at Patricia's present, these characters seem to accept a patriarchal family as the most positive outcome available, remaining "beings-for-the-others" (Lagarde *Cautiverios* 824-825) instead of being given the chance to embrace an agenda of their own. Understanding Gregoria and Patricia as part of the same genealogy solves both the Oedipal and the immigration issues. The plot, instead of tragically ending with the hanging of Jocasta or the shooting of Martina, concludes with the happy simulation of a family. Near the end of the movie, when Gregoria tells Damián: "Your family is leaving," and, a second later, "Do you want them to go?" the return to the Oedipal anomaly continues to loom in the background. It should not go unnoticed, however, that what Gregoria is actually doing is establishing for the characters as for the audience who has the authority at that moment. It is not Gregoria, it is not Patricia; it is, ultimately, Oedipus—that is, Damián.

Even though both films provide a negative characterization of Oedipal tension, it is telling how different their strategies are for solving this tension. In *Furtivos*, we can see the presence of naïve expectations regarding the audience. In *Flores de otro mundo*, the screenplay relies heavily upon genre motifs to make their message more obvious. This difference can be easily traced to the political evolution of Spain. During the dictatorship, the repeated attempts to curtail any kind of subversive message ultimately resulted in suggesting tacit agreement with the audience, equating film and propaganda beyond any scientific proof. Kathleen Vernon remind us of the overwhelming evidence against such simplistic assumptions: "[a] historical overview [...] reveals cinema's role in the process of national reinvention, and its ultimate unreliability as a vehicle for attempts to project an image of some fixed national character" (264). Polarization between the Franco regime and its dissidents prevent both of them from recognizing their shared view that film has an intrinsically "political" nature ("political" being a euphemism for "propagandist"). The principles of this issue have been

¹⁰ [La sororidad emerge como alternativa a la política que impide a las mujeres la identificación positiva de género, el reconocimiento, la agregación en sintonía y la alianza] (Lagarde "Pacto" 125).

discussed by Noam Chomsky in *Chronicles of Dissent*, where he indicates that:

One of the ways you control what people think is by creating the illusion that there is a debate going on, but making sure that that debate stays within very narrow margins. Namely, you have to make sure that both sides in the debate accept certain assumptions, and those assumptions turn out to be the propaganda system. As long as everyone accepts the propaganda system, then you have a debate. (63)

The disappearance of political censorship makes it difficult to believe in such a direct influence of film in society. As a consequence, *Flores de otro mundo* relies on more traditional genre motifs that carry their own set of expectations, which in turn may undermine other messages arising from the action. At the end of the movie, the new order is symbolized by the picture of the First Communion of one of Patricia's children. Ending on such a note serves the same purpose as a wedding or a dance. In the Spanish tradition, the finishing of a play (or movie) with a big and cheerful Catholic celebration—a wedding, a baptism, or a First Communion—is rarely symbolic of any actual social justice or, even less, gender equality. Rather, such a conclusion usually suggests the achievement of a kind of poetic justice—a resolution that audiences have enjoyed since the golden age of theater. As established by Alexander A. Parker many years ago in his seminal article about the specificity of genre motifs in the Spanish “comedia” of the Golden Age, when it comes to poetic justice, it is “required not only that the guilty should suffer but also that there should be no innocent victim” (46). While Martina pays a steep price for her transgression, the tacit pact between Gregoria and Patricia allows both women to establish a feminine genealogy of sorts. As a consequence, the former is acquitted of her punishable role in the Oedipal tension that the film portrays. As Irigaray famously established, these feminine genealogies, “are conditions for ending women's status as sacrificial objects” (Whitford 159), a point well illustrated by the peaceful end of the plot. Nevertheless, Patricia, paraphrasing Irigaray, “has been exiled in the house of her husband” (19), an environment in which she can only assert her place by dutifully accepting her role in the performance of the illusion of a family. This choice structures her experience as belonging in the same genealogy as Gregoria; however, this is quite a paradoxical way of not becoming “an accomplice in the murder of the mother” (Irigaray 19). For it is quite clear that the family that we see on the screen is not of Damián's and Patricia's making (biologically, legally, or otherwise).

On the contrary, it is a portrayal of a cultural and social construct based on a tradition in which female characters can only hope and wait, while men decide, and society acquiesces.

Works Cited

Balfour, Sebastian. "Spain from 1931 to the Present." *Spain. A History*. Ed. Raymond Carr. Oxford: Oxford U. Press, 2000. 243-282. Print.

Baudrillard, Jean. *The System of Objects*. Transl. James Benedict. James Benedict. London: Verso, 1996. Print.

Besas, Peter. *Behind the Spanish Lens. Spanish Film under Fascism and Democracy*. Denver: Arden, 1985. Print.

Bollain, Iciar and Julio Llamazares. *Cine y literatura. Reflexiones a partir de Flores de otro mundo*. Ed. Joaquín Rodríguez. Madrid: Páginas de Espuma, 2000. Print.

Borau, José Luis, and Manuel Gutiérrez Aragón. *Furtivos Guiones*. Ed. Juan J. Vázquez. Huesca: Diputación, 2009. Print.

Chomsky, Noam. *Chronicles of Dissent: Interviews with David Barsamian*. Monroe: Common Courage, 1992. Print.

Evans, Peter. "Furtivos (Borau, 1975): My Mother, My Lover." *Spanish Cinema. The Auterist Tradition*. Ed. Peter Evans. Oxford: Oxford U. Press, 2000. Print.

Fernández Heredero, Carlos. "Furtivos". *Antología crítica del cine español 1906-1995. Flor en la sombra*. Ed. Julio Pérez Perucha. Madrid: Cátedra/Filmoteca Española, 1997. Print.

Flores de otro mundo. Dir. Iciar Bollain. Int. José Sancho, Luis Tosar, Lissete Mejía, Chete Lera, Marilyn Torres, and Elena Irureta. La Iguana-Alta Films, 1999. Film.

Furtivos. Dir. José Luis Borau. Int. Ovidi Montllor, Lol Gaos, Alicia Sánchez, and José Luis Borau. El Imán, 1975. Film.

Girard, René. *Violence and the Sacred*. Baltimore: John Hopkins U. Press, 1977. Print.

Gubern, Román. *Mensajes icónicos en la cultura de masas*. Barcelona: Lumen, 1974. Print.

Irigaray, Luce. *Sexes and Genealogies*. New York: Columbia U. Press, 1993. Print.

Kinder, Marsha. *Blood Cinema. The Reconstruction of National Identity in Spain*. Berkeley CA: University of California Press, 1993. Print.

Lagarde y de los Ríos, Marcela. *Los cautiverios de las mujeres: madresposas, monjas, putas, presas y locas*. México D.F.: Universidad Nacional Autónoma, 2001. Print.

—. “Pacto entre mujeres. Sororidad.” *Aportes* 25. 123-135. Web. 21 Aug. 2011. <http://www.asociacionag.org.ar/pdfaportes/25/09.pdf>. Accessed February 22, 2012.

Marín, Karmentxu. “‘Caravana de mujeres’ para los solteros de Plan.” *El País*. 1 Jan, 1985. Web. http://elpais.com/diario/1985/01/10/espana/474159601_850215.html. Accessed February 20, 2012.

Parker, Alexander A. “The Approach to the Spanish Drama of the Golden Age.” *The Tulane Drama Review* 4.1 (September 1959): 42-59. Print.

Riambau, Esteve. “El cine español durante la Transición (1973-1975): Una asignatura pendiente”. *Un siglo de cine español*. Coord. Román Gubern. Madrid: Academia de las Artes y las Ciencias Cinematográficas de España, 2000. 179-190. Print.

—. “‘A escondidas’: una cuarta vía del cine español.” *Furtivos Borau*. Huesca: Diputación Provincial, 2009. 121-134. Print.

Sánchez Vidal, Agustín. *Borau*. Zaragoza: Caja de Ahorros, 1990. Print.

Whitford, Margaret, Ed. *The Irigaray Reader*. Oxford: Blackwell, 2004. Web. 21 Aug. 2011.

Vernon, Kathleen M. “Culture and Cinema to 1975.” *The Cambridge Companion to Modern Spanish Culture*. Ed. David T. Gies. Cambridge: Cambridge U. Press, 1999. 248-266. Print.

Politics of Display: The Museum, the Academy, and the Framing of American Indian Texts

Kim Lawson

Utah Valley University

Abstract

The representation of minority texts within the academy is an issue that is fraught with risk. The ongoing colonization of American Indian tribes within the United States further complicates the representation of Native texts. This essay examines two potential pitfalls of representing American Indian literature within the university classroom, fetishization and aestheticization, and proposes a model of translation that assists teachers and students alike in approaching Native texts in a more fruitful way.

To celebrate its 150th anniversary, the Smithsonian Institute compiled some representative pieces from its vast holdings and created a touring exhibit called “America’s Smithsonian.” This collection boasted a broad spectrum of museum pieces ranging from fine art to moon rocks and Hollywood memorabilia. In the historical section of the exhibit, entitled “Remembering,” there were a few items of American

Indian clothing, including a Ghost Dance dress labeled “Southern Arapahoe, circa 1890.” The dress, displayed as if on an invisible body, had only a brief caption below it identifying it as a particularly fine Ghost Dance dress and giving the approximate date it was made, its tribal affiliation, and a one-sentence description of the function of the Ghost Dance.¹ For me, the description raised many questions about the acquisition of the dress, about the events at Wounded Knee,² and about Native religious practices. This dress, worn by the “vanished” Indian, seemed not to demand a *remembering* of the viewer, but to reaffirm a cultural forgetfulness signaled by the absent body, the absence of context, and the all-too-brief commentary offered on the caption card.³ The presence of only very old Native artwork in the traveling exhibit, as well as the brief description that ignored the profound cultural and historical impact of the Ghost Dance up to contemporary times, reinforced

¹ The Ghost Dance message was spread by the Piute prophet Wovoka, who preached that performing the Ghost Dance, a dance honoring the dead, would result in the current world being destroyed “and replaced by a fresh one. The dead would live again and everyone would be young and happy. The buffalo would return and the white man would disappear” (Reilly 283). Wovoka’s message was peaceful, although some groups, such as the Sioux, added a strain of militarism to this prophecy. Wovoka revised and reworked the idea of the Ghost Dance that had been proposed by another Piute prophet 20 years earlier.

² The 120 men and 230 women and children in a mixed band of Sioux, following Minneconjou chief Big Foot, were fleeing the Cheyenne River reservation after Sitting Bull was killed. The U.S. military feared that the group was entering the Badlands to join hostile Ghost Dancers, so the Seventh Cavalry was sent to intercept and capture them to incarcerate them at a military prison in Omaha, Nebraska. After the Sioux were surrounded, Colonel Forsyth demanded that they give up their guns. The Sioux were hesitant to do this after Sitting Bull had been shot to death while unarmed. One Sioux man defiantly stated that he had paid good money for his gun and would not give it up. A scuffle ensued over his gun, and when it went off, the two sides began shooting at each other. The U.S. military had the superior firepower, with 470 soldiers and 4 Hotchkiss guns (small cannons), and the Sioux began to flee in all directions. They were quickly chased down and killed. “In the end, nearly 300 Indians were killed, mostly women and children. Sixty soldiers, many of whom were victims of ‘friendly fire,’ were also killed” (Reilly 284).

³ While museums routinely display clothing items on hidden wire forms instead of on mannequins, to display a Ghost Dance dress in this manner seems to support the popular 19th and 20th century stereotype of the Vanishing Indian that lays the groundwork for the erasure of American Indians from the cultural consciousness. The Wounded Knee Massacre is remembered as a final military and spiritual resistance of the Sioux to the U.S. Government’s policy of relocation. After the 1890s, the Indian Problem was “contained.”

the myth of the vanished but noble race of Indians. This forgetting, (mis)labeled as remembering, echoes Homi Bhabha's discussion of "the historical memory of the Western nation which is 'obliged to forget'" (165). He states that

it is through this syntax of forgetting—or being obliged to forget—that the problematic identification of a national people becomes visible.... To be obliged to forget—in the construction of the national present—is not a question of historical memory; it is the construction of a discourse on society that performs the problem of totalizing the people and unifying the national will.... Being obliged to forget becomes the basis for remembering the nation, peopling it anew, imagining the possibility of other contending and liberating forms of cultural identification. (160-1)

The space created by this willful cultural amnesia and absence of historical information allows museum viewers to fill in the blanks as they wish concerning Native histories and cultural practices and to bring to their viewing the colonialist assumptions that circulate within the larger culture.

Witness the conversation I overheard in front of the Ghost Dance dress. One woman commented to another that it was striking that the dress seemed to be patterned after the American flag with its red, white, and blue color theme and star pattern. The second woman answered back that it was indeed ironic to show allegiance to Old Glory as a Ghost Dancer considering that nearly 300 dancers were killed by the U.S. military at Wounded Knee alone. While I was surprised that one of these viewers was able to connect this display with Wounded Knee (no connection was made for the audience in the display itself), I did also notice that these two women viewed the dress from the lens of their own national and cultural matrix of meaning, which seemed to force a connection between the dress and the American flag that wasn't there. The second woman's ethnocentrism caused her to deploy the facts she knew about Wounded Knee to arrive at distorted conclusions. This dress had a painted design on white canvas that used blue and red paints, blue on the top or bodice of the dress and red on the skirt, and the paint left white spaces in the shape of stars. The shapes of animals, such as magpies, a turtle, and a rabbit with a person in the center, were also painted on the dress, but these designs seemed invisible to the women who were examining the dress: They saw only stars and stripes, though there were no stripes present. The problem illustrated by the two women viewing this dress was not with the viewing of this particu-

lar piece but with the concept of display in general when objects are displayed in a near-vacuum of context. This decontextualization is a mark of colonization, which has serious real-world effects such as the inability of these two women to see the cultural specificity of the dress but instead to import it wholesale into their own meaning-making system—in effect, to coopt a Native religious object into a symbol of U.S. nationalism. The example of the Ghost Dance dress should suggest from the mode of display of the item, as well as by the reading offered by the two viewers of the display, that “museums and the museumizing imagination are both profoundly political” (Anderson 178).

In “The Post Always Rings Twice: The Postmodern and the Post-colonial,” Linda Hutcheon compares the museum to the academy: “both could be said to work toward the acquisition of knowledge through collecting, ordering, preserving, and displaying—in their different ways—the ‘objects’ of human civilization in all its varieties” (206). Although Hutcheon does not elaborate at length on this comparison, it does seem to be a fruitful line of inquiry. The academy, like the museum, is a place where people go to learn things that they don’t know, to “read” what is displayed in the museum or in the college curriculum, and to use this information to help them make decisions about “value” and “truth.” But what messages are being communicated to audiences about the objects on display? And who composes this audience?

With regard to the presentation of American Indian literature in the academy, there are two forms of display that I will critique in order to tease out their shortcomings and offer alternative strategies for the presentation of American Indian literature. These two different but related forms of display are fetishization⁴ and aestheticization. Similar categories have been proposed in the study of museum display, and Ivan Karp, examining various museum exhibitions of marginalized art, echoes a useful and commonly made division in his discussion of museum display:

An exhibiting strategy in which difference predominates I call *exoticizing*, and one that highlight similarities I call *assimilating*. We are more familiar with exoticizing strategies; they predominate in popular-culture forums such as travel accounts, as well as in some academic writing. Assimilating

⁴ Emily Apter, in the Introduction to *Fetishism as Cultural Discourse*, offers several analyses of fetishism, one of which is the “smoking out [of] the Eurocentric voyeurism of ‘other-collecting’” (3).

strategies are less easy to read. They appeal to the audience's sense of the familiar and natural. They don't stop exhibition-goers in their tracks and produce a 'what in the world is that?' response. (375-6)

Karp's distinctions seem to run parallel to the methods of display of American Indian literature that I will discuss, his category of the exotic aligning with what I refer to either as fetishization or exoticization, and his strategy of assimilation aligning with what I refer to as aestheticization. As Karp points out later in the same passage, the aesthetic imposed on the displayed object is the aesthetic of the dominant culture, through which the object is evaluated.

When an American Indian text is "displayed" in a literature course, it is possible for some well-meaning professors to turn the text into an *objet d'art* and ignore issues such as cultural specificity or political and ideological underpinnings of a given text. I have even heard myself making arguments to a class that Native artistic productions, both ancient and contemporary, can *also* compete in the arena of fine art. This assertion alone often requires significant argument on the part of the teacher. Additionally, some feminists may misread or misrepresent American Indian women's texts or may address the texts only as "women's writing," thus ignoring the cultural and political specificity of these texts.⁵ I believe that this move is always dangerous, but it is particularly so with Native texts because of the urgent, contemporary issues that Native nations face, such as legislated identity, national sovereignty, and stereotypic and demeaning representation. These issues are not merely trivialized, but can be completely erased in some classrooms where American Indian texts are read. This aestheticization of American Indian writing is a colonizing move. American Indian texts are often treated as just another "object," in Hutcheon's terms, in the museum of the academy, and this decontextualization leads to recolonization of American Indians on a different, newly developed front marked by voyeurism and consumption. Critics settle the territory of the text using some of the same mentalities which allowed Europeans to claim Native lands without remorse, such as the alleged superior

⁵ Much interesting and important scholarship is being done at the intersection of feminist studies and minority/ethnic studies surrounding questions of essentialism. For some studies in this area, see articles by Uma Narayan, Seyla Benhabib, Nancy Fraser, Norma Alarcon, Maxine Baca Zinn, and Bonnie Thornton Dill.

ability to develop it and the pretended absence of any Indians in the territory.

Even a perusal of how Native texts are treated in academic literary journals is highly instructive of the broad range of methodologies used to analyze American Indian literature. There are certainly many excellent treatments of Native texts springing from a wide range of theoretical stances. There are also, however, some articles in which the issue of Indian identity is somehow quietly dismissed before the “real work” of literary analysis begins. As Paula Gunn Allen states, “by using only the techniques of explication, without seeking to understand the context from which materials derive, the critic has removed the book and its author from the living web of the people and tradition from which they both arose” (xi). While considering the culture from which a text comes is by no means an idea new to the academy or to literary analysis, ignoring the colonial implications of rendering American Indian cultures invisible or trivializing difference in Indian life ways and ideologies is to perpetuate the mindset that supports the continued exploitation and colonization of American Indian cultures and land base.

Another possible conclusion drawn from the aestheticizing approach is that American Indian literature simply cannot measure up to other bodies of literature. The belief that American Indian literature is technically or artistically inferior to texts in the Western tradition is still a common belief, even at the college level. An illustration of this belief occurred in a graduate-level course I took in which I had to defend my desire to write on Leslie Marmon Silko’s *Ceremony* against charges that the text might not be dense or rich enough at the level of a “close reading” to successfully fulfill the course requirements for the term paper. The professor, incidentally the highest paid professor in the department at the time, implying that he was the finest scholar, seemed to be functioning within colonialist stereotypes about the quality and value of Western art vis-à-vis non-Western art and making judgments about American Indian texts being too shallow or lightweight for the rigorous work he wanted to perform within a Western philosophy of literary theory and analysis. Silko’s text seemed to resist the type of reading that the professor wanted me to perform in a way that other texts may not have. During my work-in-progress report, some other students as well as the professor suggested that I analyze the book in terms of how it engages a modernist aesthetic; however, this approach seemed to me an attempt to justify Silko’s artistic merit by attempting to compare her favorably with the Western artistic values exhibited in modernism. This is not to say that those artistic practices do not appear in Silko’s novel or that she remains uninfluenced by modernist novelistic and artistic conventions. The fact that she has written novels sug-

gests that she has entered into a genre with strong Western European historical traditions and is aware of the genre conventions and probably their origins, as well. Certainly the novel is not a form indigenous to the Laguna Pueblo. My intent is rather to point out that it did not occur to this group of academics that Silko's novel could be analyzed, explicated, or evaluated on its own terms through an indigenous aesthetic.⁶

This personal example illustrates a problem in the treatment of American Indian literature that also exists in the treatment of other postcolonial literature such as African literature. Anthony Appiah criticizes "the post-colonial legacy which requires us to show that African literature is worthy of study precisely (but only) because it is fundamentally the same as European literature" (Appiah qtd. in Gates 14). In fact, some teachers who integrate minority texts into their curriculum attempt to validate these texts by illustrating through lengthy analysis that these texts meet the generations-old Western aesthetic criteria used to evaluate literary art within the academy and support systems for this type of reading are firmly in place. This seems to be a subtly paternalistic, colonizing move within the academy with regards to Native texts, because it maps the terrain of evaluation of these texts onto a Western aesthetic. This move ignores issues of cultural specificity, historicity, and ideologies that should inform academic evaluations of texts.

The flip side of the coin of aestheticization of Native artistic productions is exoticization and fetishization. This logic creates an interesting spin on the potential rejection of American Indian literature based on failure to meet aesthetic standards because the assignment of value based on exoticism creates an almost "back-door" approach to the admission of American Indian literature into the academy and the classroom. Ivan Karp asserts that in the museum, the display of the exotic serves to show "how a well-known practice takes an inverted form among other peoples," and this same mechanism of display may be deployed in a classroom to construct American Indian literature as a foil or counter-example for Euroamerican literature (376). While there is certainly nothing wrong with noticing differences in different bodies

⁶ Steven Leuthold, in *Indigenous Aesthetics: Native Art, Media, and Identity*, explains that the term indigenous aesthetic "refers to the complexities and contradictions found in the art and media of indigenous peoples today and the ways that their aesthetic experiences inform, enrich, and challenge members of non-native cultures" (ix). He further claims that "the meanings of concepts such as 'art' and 'aesthetic' may be extended, challenged, and even subverted through the study of indigenous ideas and expressions" (x). One wonders what would have happened in that graduate course if I'd had the presence of mind to suggest that we should examine a modernist novel through an indigenous aesthetic.

of literature, focusing primarily on how American Indian literature is of interest simply because it *differs from* mainstream American literature is insulting because it again creates value out of how American Indian literature compares against Euroamerican literature, only this time valorizing the differences instead of the similarities, as with an aesthetic comparison. One respected literature anthology provides an example of the vacuous non-logic of fetishization in the teacher's resource guide, which points out that Silko's short story, "'Yellow Woman' is built on different traditions from those in the cultural background of most American students" (Charters and Charters). This statement seems to suggest to instructors that pointing out difference alone is quite enough to suggest the validity of presenting the story to their class. The absence in the teacher's resource guide of any discussion of what some cultural differences might be, how they affect the reading of the story, or what these differences reveal about the two cultures seems to suggest that questions like these are not vital to the study of the literatures of other cultures.

Marianna Torgovnick, in *Gone Primitive: Savage Intellectuals, Modern Lives*, discusses the politics of displaying the art and artifacts of so-called primitive cultures. She asserts that

within the dominant narrative as told by art historians, the 'elevation' of primitive objects into art is often implicitly seen as the aesthetic equivalent of decolonization, as bringing Others into the 'mainstream' in a way that ethnographic studies, by their very nature, could not. Yet that 'elevation' in a sense reproduces, in the aesthetic realm, the dynamics of colonialism, since Western standards control the flow of 'mainstream' and can bestow or withhold the label 'art.' (82)

This statement could be made to apply to the study of literature and reveals some of the prejudices of the discipline in that the inclusion of American Indian texts on syllabi and in critical work is often interpreted as bestowing the label of art onto texts, and the inclusion itself should be a satisfactory end. The "elevation," in Torgovnick's words, of American Indian texts to the status of literary art that merits critical attention should not be interpreted as a desirable end in itself, especially if critics and teachers go no further in their discussion and analysis of these texts than applying Western European tools that have so often yielded unflattering or derogatory interpretations of American Indian literature vis-à-vis Western literature and literary art.

An alternative model is needed for the teaching of American Indian literature to help instructors avoid the pitfalls just discussed, and

the model I propose is one centered on the practice of translation. Karp (382) asks, "How do styles of exhibiting position exhibitors and audiences vis-à-vis the people of the cultures exhibited?" This question surrounding styles of exhibition seems to highlight the issue of mediation and recognizes that when one displays art from other cultures in one's own culture, one necessarily and unavoidably mediates between cultures. This mediating, or meaning-making, is nothing short of translation. The issue of translation is not given enough attention currently by teachers who integrate American Indian texts into their classroom. Translation is frequently associated exclusively with the idea of translating texts from one language into another, but the translation of ideologies, cultures, or belief systems into another culture is also an issue of translation, and as such, should be treated as a vital endeavor. "A work is, in a sense, 'translated,' that is, displaced, transported, carried across, even when it is read in its original language by someone who belongs to another country and another culture or to another discipline" (Miller 207). If we accept Miller's assertion, then we must concede that even contemporary literature by American Indians demands attention to translation issues because non-Native students are reading across cultures. The presentation and critical analysis provided by the instructor needs to make the illumination of the text to the students its main focus, not the obsessive fixation on cultural differences as an exotic source of titillation; however, a balance of focused and respectful study is often elusive. Wolfgang Iser claims that there is an "ethics inherent in cross-cultural discourse" because of the fact that "a cross-cultural discourse requires a certain amount of self-effacement, perhaps a suspension of one's own stance, at least for a certain time, in order to listen to what the others are trying to say" (302). This self-effacement is a move that is vital, even as it is fraught with risk. Gayatri Spivak juxtaposes risk against safety in her discussion of translation, with the category of safety indicating a poor translation in as much as it does not account for anything but grammatical, linguistic translation. Alternatively, translation that engages risk also considers the rhetorical nature of language and the rhetorical inferences that permeate languages *Outside in the Teaching Machine* 180). Thus, translation should not be a unidirectional performance of authority and expertise but should remain in dialog with both the text and the culture from which it emerged. We should seek to create a politics of translation that does not repress "translation as dialogue" and does not reduce translation to a monologue (Cheyfitz xxv).

Greg Sarris, in *Keeping Slug Woman Alive: A Holistic Approach to American Indian Texts*, proposes a model for cultural translation that constructs a continuous dialogue between readers, writers, students, and

teachers in which “each group can inform and be informed by the other” (7). He asserts that for most readers of American Indian literature, reading is necessarily cross-cultural and that students should be taught to interrogate not only the texts they examine, but also the way they read the texts and why they read the texts in a certain way. Students should be pressed to consider their own personal and cultural boundaries along with the textual boundaries of the literary productions to better understand, through continuing dialog and inquiry, both themselves and that which is other. This exchange between self and other is interrogated to increase awareness of one’s current beliefs, positions, and cultures. Students of American Indian literature should be asked to consciously contextualize themselves, their reading practices, their ideological positions, and the texts that they examine. American Indian literature can create for non-Native readers “a dialectical or two-sided journey examining the realities of both sides of cultural differences so that they may mutually question each other, and thereby generate a realistic image of human possibilities and a self-confidence for the explorer grounded in comparative understanding rather than ethnocentrism” (Fischer qtd. in Ruppert viii). This exercise of mediative scholarship is capable of propelling the reader into the fertile but largely uncharted terrain of cross-cultural exchange.

Enacting Sarris’s pedagogical suggestions for cultural translation in a classroom setting is a difficult undertaking. Engaging students in a critique of their current positions vis-à-vis the text is an ambitious pedagogical task. I have found that the Yellow Woman stories in Silko’s *Storyteller* work well as springboards both to help students interrogate the reading practices and ideologies that they brought to texts and to engage students in productive cultural translation that helps them interpret the literature within and through its cultural context. My students were initially confused and offended by this series of stories, which, for them, seemed to emerge from rape fantasies or self-indulgent escapism of the female protagonist. The Yellow Woman myth that Silko is referring to tells of Kochininako, a Pueblo woman, who meets Buffalo Man on a river bank one day. She agrees to leave her home and family to join him because he is sexually irresistible to her (Silko 54-62). In the traditional stories, Yellow Woman usually returns to her home and family later. My students felt that the sexual dynamic was “disturbing” (their word choice) and postulated that perhaps Yellow Woman was a martyr figure or that she may occupy a degraded position because of her lack of resolve. Next, the students attempted to perform a Western feminist reading of Yellow Woman, interpreting the stories not as describing a weak female but as stories about fulfilling personal longings,

about leaving one's community to experience new sensations, and about the rejection of social conventions.

For that same week, I had also assigned my students to read an interview with Silko in which she explains the function of Yellow Woman in Laguna Pueblo culture and how Yellow Woman's sexual appetites could be interpreted through a different lens that would not cast her in a negative light. In this interview, Silko provides cultural clues on how to read her stories from within the culture from which they emerged rather than through the lens of Euroamerican mainstream culture. Silko explains that the sexual attraction that Buffalo Man and Yellow Woman feel toward each other symbolizes a continuing reconnection between the human world and the animal and spirit worlds. She states that traditional Yellow Woman stories are regenerative stories, and the powerful sexual drives in the story illustrate the intimate and necessary connection between humans, animals, and the spirit world. Silko states, "The link is sealed with sexual intimacy, which is emblematic of that joining of two worlds" (Barnes 95). Traditional Yellow Woman stories often note the benefits that the Laguna Pueblo would experience after the return of Yellow Woman, illustrating the benefits that the community reaped through the symbolic union of Yellow Woman and Buffalo Man (Ruoff 72). In the interview, Silko explains that the Laguna Pueblo are a matrilineal people, with their houses and land passed down through female lineage. Figuring a female as representing humanity within a human world/spirit world dichotomy may illustrate to students the central role of women within the Laguna Pueblo culture.

The previous example illuminates several of Sarris's ideas about teaching American Indian texts and about the cultural translation that needs to take place when students read cross-culturally. These students learned through experience that cultural biases influence interpretive acts and that students can transform the problems that they have in understanding a text into the basis for their research and a starting point for their dialog with both the text and the culture from which it emerges. As the teacher, I interceded after the students had discussed their reactions to the Yellow Woman stories, then had attempted to interpret the stories within their own culture, and finally, had begun to wonder what these stories might mean within Laguna Pueblo culture, but had inadequately comprehended Silko's explanation of Yellow Woman's function within Laguna Pueblo culture. The students were struck by a woman's sexual desire being cast as a positive part of a cultural mythology and wondered about the way women's sexuality is figured in their own culture. This question led students to offer observations about the roles of women in different societies based on their perceived worth

or centrality to the culture and also about the relationship between humans and the natural world. These students were now at least partially able to examine what the two different interpretations of the Yellow Woman stories and different figurations of female sexuality in a story might reveal about the very different backgrounds that had come together in the reading of this text.

It is at this point in their intellectual progress that these students were attempting a cultural translation because they were trying to move past their default interpretive lenses to see both themselves and the text from a new perspective in their effort to make meaning from the text. This attempt at cultural translation is perhaps more like a border crossing than a seamless translation of cultural constructs from one culture into another because one cannot ever translate the other into the same. These students experienced the limits of translatability in as much as the experience of untranslatability is the education that they gained. Laguna Pueblo constructions of femaleness, of balance maintained in nature, and of spirituality cannot be translated into Euroamerican mainstream culture because there is no cultural equivalent for these Native concepts, but this knowledge alone is invaluable to students of the literatures of other cultures.

Such a practice of rigorous cultural translation can provide readers with a framework within which to rethink the imperialist mindset that has governed the treatment and interpretation of American Indian artistic productions for so long. Arnold Krupat explains the pressing need to create a model of translation for contemporary American Indian literature that takes into account the “condition of ongoing colonialism” in which that body of literature is produced (32). He proposes a category that he calls “anti-imperial translation,” which he claims can function as a model

for conceptualizing the tensions and differences between contemporary Native American fiction and the “imperial center.” Because historically specifiable acts of translative violence marked the European colonization of the Americas from Columbus to the present, it seems to me particularly important to reappropriate the concept of translation for contemporary Native American literature. To do so is not to deny the relationship of this literature to the postcolonial literatures of the world but, rather, to attempt to specify a particular modality for that relationship. (32)

Krupat argues that despite the fact that contemporary American Indian writers write in English and function largely with genre con-

straints of Western literary forms, that their Native tongue, their Native world view, exerts a powerful influence on their writing, which in turn creates a “foreignness” that serves to challenge not only forms of writing in English but also cultural values and assumptions of mainstream American readers. Any reading of an American Indian text as an anti-imperial translation must demonstrate “*how* that text incorporates alternate strategies, indigenous perspectives, or language uses that, literally or figuratively, make its ‘English’ on the page a translation in which traces of the ‘foreign tongue,’ the ‘Indian,’ can be discerned” (38). Silko’s *Yellow Woman* stories perform this type of translation, as a brief reexamination of the Silko’s text illustrates.

Silko includes in her collection *Storyteller* a lengthy narrative poem that attempts to recreate a traditional story about *Yellow Woman*, but she also includes another short story that literally translates the traditional story of *Yellow Woman* into a present-day event recounted by a narrator and that conforms to the genre conventions of a contemporary short story. This short story resists the ideological pull of the imperial center by not only adhering to, but also promoting, traditional Laguna Pueblo beliefs through integrating them into contemporary events. Laguna Pueblo ideas about the spirit world dominate the short story and the recounting of a meeting between a human and a spirit person are treated as completely plausible events despite the present-day setting. Contemporary characters reenact sacred events in the latest of many incarnations of an ongoing spiritual process. Although Euroamericans have attempted to intervene in Native spiritual practices and beliefs for centuries to convert American Indians to the dual religions of Christianity and Western science, Silko enacts a resistance to the imperial center by representing the successful integration of traditional beliefs and indigenous perspectives into contemporary life through this anti-imperial translation. Her importation of a traditional, orally transmitted sacred story into a Western European literary genre is a profound and daring revision, where the term revision, “particularly in ‘postcolonial’ critical contexts, suggests counternarrative or revisionist history” (Cox 5).

Silko’s text is only one example of an American Indian text that affords opportunities for cultural translation. Excellent work in cultural translation is currently being done in many university classrooms with pedagogy being figured as a contested political site. The performance of cultural translation certainly has political ramifications because of its ability to disrupt racial and ethnic stereotypes as well as to provide a base for cross-cultural exchange that can influence the way students think about American Indian cultures. Spivak claims that “in order to recognize [teachers and students across the country who are attempting

to keep alive a critical cultural practice], pedagogy as political interpretation must be seriously considered" *In Other Worlds* 124). The study of American Indian texts within college curriculums has the potential to create opportunities for cultural translation, but as Lynet Uttal points out, "diverse inclusion is not enough if these token voices have no impact and influence on all our ways of thinking" (45). The academy, much like the museum, shapes, through its classifications and representations, "how and what we learn of the Other" (McLoughlin 4). Vigilant, responsible instructors must include American Indian texts in our courses and curriculums; however, we must examine them on their own terms, taking their cultural specificity into consideration, and we must realize that the teacher, as mediator, needs to present cultural translation as a necessary methodology for cross-cultural reading.

Works Cited

Anderson, Benedict. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Rev. ed. New York: Verso, 1991.

Apter, Emily. "Introduction." *Fetishism as Cultural Discourse*, ed. Emily Apter and William Pietz. Ithaca: Cornell U Press, 1993. 1-9.

Barnes, Kim. "A Leslie Marmon Silko Interview." *The Journal of Ethnic Studies* 13.4: 1986. 83-105.

Bhabha, Homi K. *The Location of Culture*. New York: Routledge, 1994.

Charters, Ann, and Samuel Charters, eds. *Resources for Teaching Literature and Its Writers: An Introduction to Fiction, Poetry, and Drama*. Boston: Bedford Books, 1997.

Cheyfitz, Eric. *The Poetics of Imperialism: Translation and Colonization from The Tempest to Tarzan*. Rev. ed. Philadelphia: U of Pennsylvania Press, 1997.

Cox, James H. *Muting White Noise: Native American and European American Novel Traditions*. Norman: U of Oklahoma Press, 2006.

Gates, Henry Louis, Jr. "Editor's Introduction: Writing 'Race' and the Difference it Makes." *Race, Writing, and Difference*, ed. Henry Louis Gates, Jr. Chicago: U of Chicago Press, 1986. 1-20.

Hutcheon, Linda. "The Post Always Rings Twice: The Postmodern and the Postcolonial." *Textual Practice* 8.2: 1994. 205-38.

Iser, Wolfgang. "Coda to the Discussion." *The Translatability of Cultures: Figurations of the Space Between*, ed. Sanford Budick and Wolfgang Iser. Stanford: Stanford U Press, 1996. 294-302.

Karp, Ivan. "Other Cultures in Museum Perspective." *Exhibiting Cultures: The Poetics and Politics of Museum Display*, ed. Ivan Karp and Steven Lavine. Washington: Smithsonian Institution Press, 1991. 373-85.

Krupat, Arnold. *The Turn to the Native: Studies in Criticism and Culture*. Lincoln: U of Nebraska Press, 1996.

Leuthold, Steven. *Indigenous Aesthetics: Native Art, Media, and Identity*. Austin: U of Texas Press, 1998.

McLoughlin, Moira. *Museums and the Representation of Native Canadians: Negotiating the Borders of Culture*. New York: Garland, 1999.

Miller, J. Hillis. "Border Crossings, Translating Theory: Ruth." *The Translatability of Cultures: Figurations of the Space Between*, ed. Sanford Budick and Wolfgang Iser. Stanford: Stanford UP, 1996. 207-23.

Reilly, Hugh J. "Wounded Knee, South Dakota, Massacre at." *Encyclopedia of American Indian History*, ed. Bruce E. Johansen and Barry M. Pritzker, vol. 1 (Santa Barbara: ABC-CLIO, 2008). 282-5.

Ruoff, LaVonne Brown. "Ritual and Renewal: Keres Traditions in Leslie Silko's 'Yellow Woman.'" *"Yellow Woman:" Leslie Marmon Silko*, ed. Melody Graulich. New Brunswick: Rutgers U Press, 1993. 69-81.

Ruppert, James. *Mediation in Contemporary Native American Fiction*. Norman: U of Oklahoma Press, 1995.

Sarris, Greg. *Keeping Slug Woman Alive: A Holistic Approach to American Indian Texts*. Berkeley: U of California Press, 1993.

Silko, Leslie Marmon. *Storyteller*. New York: Arcade Publishing, 1981.

Spivak, Gayatri Chakravorty. *In Other Worlds: Essays in Cultural Politics*. New York: Routledge, 1988.

Spivak, Gayatri Chakravorty. *Outside in the Teaching Machine*. New York: Routledge, 1993.

Torgovnick, Marianna. *Gone Primitive: Savage Intellects, Modern Lives*. Chicago: U of Chicago Press, 1990.

Uttal, Lynet. "Inclusion Without Influence: The Continuing Tokenism of Women of Color." *Making Face, Making Soul: Haciendo Caras. Creative and Critical Perspectives by Feminists of Color*, ed. Gloria Anzaldua. San Francisco: Aunt Lute Books, 1990.

Jules Verne's *The Castle of the Carpathians*: Gothic Castles and Science Fiction in a Literary Landscape

Ace G. Pilkington

Dixie State College

Abstract

The Castle of the Carpathians has received less attention than it deserves. It is a Gothic love story that has puzzled people who expected something like Verne's early work. This story suggests that different people see the same events in different ways and that even the same people see the same events differently if their perspectives or technologies change. At the center, is La Stilla, a beautiful opera singer pursued by two men. Her death destabilizes them both. She serves many purposes in the novel; one of them is to demonstrate the dangers of fame. La Stilla also may be the embodiment of the author's own fears. Verne knew what it was like to face adulation and how dangerous passions could be. If in this work he foresaw fanatical admirers struggling over what was left of his voice, he could claim one more accurate prediction.

The Castle of the Carpathians is a good novel by a great writer, and as usually happens with such works—Shakespeare's *King John* or *Cymbeline*, for instance—it has received far less attention than it deserves. The sheer size and the unique shape of Jules Verne's reputation both militate against fair treatment. However, Jules Verne is currently under attack or perhaps, more accurately, he has become a battleground in the continuing wars of literary critics, and out of that plentiful shedding of ink and wounding of reputations may well come just the attention that *The Castle of the Carpathians* ought to have. Ironically, a novel about (among many other things) the intense dangers of celebrity may benefit from Verne's own star status and the ferocious turf wars it has generated. Jules Verne worked hard to create the image and achieve the fame that gave him financial control of his life and, eventually, editorial control of his work. But his success cost him the privacy he treasured, the safety he had taken for granted, and the literary recognition for which he hungered. By the time he came to write in *The Castle of the Carpathians* about La Stilla, an artist of transcendent ability caught between two fiercely possessive admirers, Verne knew firsthand what it meant to be an object of adoration who was pulled and pummeled by his fans. And just as with La Stilla, Verne's death has quickened rather than quieted the strife.

The pre-eminent candidate for the title of first science fiction (sf) writer, first writer of techno-thrillers, and so on and on, Verne is often credited with inventing science fiction. Charles Sheffield, one of the best of twentieth-century science fiction writers and one of the few who also worked as a scientist throughout his life, presented a more realistic hypothesis. He identified two "origin points" for modern sf in his book *Borderlands of Science: How to Think Like a Scientist and Write Science Fiction*. They were Mary Shelley and Jules Verne (13-14).

Any such assertions, however qualified, must, of course, be controversial, and these are currently some of the most hotly contested claims among Verne critics, especially those who dislike and despise science fiction altogether or prefer one form of it to another. Thus, in the words of Marc Angenot, "Present-day critics are studying Verne's imaginative gifts, narrative techniques, and world view, passing over the illusory scientific or parascientific value of the novels, which has at last come to be considered simply irrelevant" (34). Even for those who are happy to discuss science fiction, there can still be questions about Verne's status as sf writer and founder. Arthur B. Evans writes in his *Jules Verne Rediscovered*, "A more likely candidate for this particular honor would be H. G. Wells, if one were to examine the thematic and structural character of the two writers' works. On the other hand, Jules Verne might quite justifiably be termed the 'father of scientific fiction.'"

The distinction is paramount. Popularizing science *through* fiction, Verne's novels contain little of what the general reading public sees as standard SF fare. One finds, for example, no E.T.s or BEMs (bug-eyed monsters) in Verne's narratives" (2).

As John F. Kennedy famously said, "There's an old saying that victory has a hundred fathers," (Beschloss 130) and I would add that science fiction has quite a few mothers as well, but clearly what is at stake here as much as anything is a definition of science fiction itself. The term did not come into popular use until about 1929 and is most often credited to Hugo Gernsback, another person who is sometimes called the "father of science fiction." He employed it, or at least the earlier form "scientifiction," frequently in his magazines; however, the *Oxford English Dictionary* finds an isolated use (by W. Wilson) as early as 1851. Most definitions of sf are argumentative at best and abusive at worst and can often be boiled down to the message that what I write (or read) is science fiction, while what you write (or read) is some sort of amorphous glop left over from a particularly repulsive evolutionary dead end.

Isaac Asimov's definition seems to me to be as sensible as any (and far more sensible than most). He says (in *Asimov on Science Fiction*) that sf is about "events played against social backgrounds that do not exist today" but "could, conceivably, be derived from our own by appropriate changes in the level of science and technology." He goes on to argue "that the field can scarcely have existed in its true sense until the time came when the concept of social change through alterations in the level of science and technology had been evolved in the first place" (17-18). So, to do a bit more boiling down, science fiction is about social change and an extrapolation from the science of the present to a possible shape for the future. Or in Jules Verne's words, "I merely use my imagination and literary skill to argue from what is possible to what may be possible—tomorrow" ("Jules Verne's Autobiography: A Collage of Interviews" in *The Jules Verne Encyclopedia*, Taves 53).

This definition would include American genre science fiction, much that was written in the nineteenth century (including Mary Shelley's *Frankenstein*), and the scientific romances of H. G. Wells and Arthur Conan Doyle (plus, of course, Jules Verne himself, though it would not grant him unique status). It would exclude everything before the scientific method became a commonplace and the speed of social change escalated to the point where it could be perceived in the course of a single lifetime.

Charles Sheffield, who always pronounced it "science fiction," with a noticeable stress on "science," found metaphor helpful in his

definition, “*Science fiction* consists of stories set on the shore or out in the shallow coastal water of that huge scientific land mass. Stay inland, safe above high tide, and your story will be not science fiction, but fiction about science. Stray too far, out of sight of land, and you are in danger of writing fantasy—even if you think it’s science fiction” (*The Borderlands of Science: How to Think Like a Scientist and Write Science Fiction* 7). That is, I think, similar to the distinction that Arthur Evans makes. Incidentally, for those who don’t like Asimov’s definition or mine or Charles Sheffield’s, here is the *OED*’s: “Imaginative fiction based on postulated scientific discoveries or spectacular environmental changes freq. set in the future or on other planets and involving space or time travel” (*The Compact Edition of the Oxford English Dictionary*, Volume III, 1039). Here’s one more, the shortest yet, from *The Star Trek Encyclopedia: A Reference Guide to the Future*, “Branch of literature and media dealing with the effects of science and technology on society” (658). If Arthur Evans is right about “what the general public sees as standard science fiction fare” (and he is a perceptive critic), then it would appear that sf writers have a different definition of science fiction than some of their readers (or viewers) do.

Certainly, Jules Verne’s work can be characterized in many ways and fitted easily into many genres. Critical objections to his status as science fiction founder or, for that matter, writer are, in part, preparation for wider-ranging studies of everything he wrote and part of the ongoing struggle to claim and, to some small extent, to control what he wrote. Even the centerpiece of his science fiction reputation, his ability to extrapolate the shape of tomorrow, is under attack. Timothy Unwin says, “Verne’s focus is almost exclusively on what is already known, and his novels are nourished above all by existing scientific, geographical, and historical documents” (6). Yes, but successful extrapolation starts with facts about the present. It is also true that Jules Verne made mistakes about science and that many of his predictions (if we insist on labeling the technological changes in his stories as such) turned out to be wrong. However, the changes he showed in his fictions often turned out to be right either because he could see what was coming or because he inspired others to bring it into being.

Although Verne did not believe that scientific change would necessarily be change for the better, there is little doubt that he epitomizes the science fiction writer as predictor and even maker of the future. “Verne’s work directly influenced scientists—not of his own generation but in the one that followed.... We know for a fact that Tsiolkovsky, the father of the Russian space program, was inspired by Verne. Hermann Oberth, whose work in turn inspired Wernher von Braun,

discovered Verne's *From the Earth to the Moon* when he was eleven years old" (Sheffield 14-15). Successful science fiction (no matter how we define it) frequently does this—whether it's H. G. Wells influencing Robert Goddard or Isaac Asimov convincing Joseph Engelberger to build the first industrial robot or even *Star Trek* suggesting work on cell phones and ion drives. However inconvenient it is to admit that Jules Verne, in the midst of his other achievements and against the weight of the great majority of his books, was a very important science fiction writer, the evidence is there.

In Jules Verne's *Paris in the 20th Century* (an interpretation of the impact of science so negative that it remained unpublished for over a hundred years), "There are electric lights in profusion; boulevards and department stores lit as brightly as the sun; gigantic hotels; great avenues filled with horseless carriages powered by internal combustion ... public transport provided by street cars and automatic driverless trains; majestic mansions fitted with elevators and electric buttons that open doors; financial hives equipped with copiers, calculators, and fax machines" (xii). So says Eugen Weber in his Introduction to the first English translation. He continues, "Paris in the twentieth century—and more specifically in 1960—teems with prodigies that were hard to imagine a hundred years before; yet Jules Verne imagined them because the science and technology of his day suggested their possibility. Imagination is the capacity to rearrange available data or to extrapolate from them, and Verne was a masterful extrapolator" (xiii).

Surprisingly often, Jules Verne's science fiction grew into science fact. Perhaps no one else has seen further off or more accurately the shape of things to come, and certainly few have had more eminent disciples to make facts from their fictions than Jules Verne did. And unlike Nostradamus and other "seers" whose vague maunderings could mean almost anything, Verne provided details so concrete as to serve, on occasion, almost as instruction manuals. I do not mean to suggest that he created new equations or engineering diagrams but only that again and again he produced fictional inventions that scientists and engineers could make in the real world—once they had come up with the equations and created the engineering diagrams for themselves.

In *Five Weeks in a Balloon* (1863), Verne correctly predicted that the source of the Nile would be found in Lake Victoria. Verne foresaw that even when a large storm was raging on the surface of the ocean, the depths would remain calm, a prediction that Simon Lake, inventor of "the first submarine to operate successfully in the open sea" (Introduction to *20,000 Leagues Under the Sea*, translated by Walter James Miller and Frederick Paul Walter, x), verified 25 years later, saving himself and his vessel because he had read and remembered

Verne's *Twenty Thousand Leagues Under the Sea*. Lake called Verne "the director-general of my life" (x). Verne was the first to see that it was possible to cross the South Pole by sailing under it (ix), and the U.S. nuclear submarine that was the first to execute such a maneuver at the North Pole was named the *Nautilus* to honor (among others) Jules Verne and his fictional submarine (ix). "Admiral Richard Byrd, on his way to the South Pole," also called Verne his "main inspiration" (x).

In *Twenty Thousand Leagues*, Verne predicted scuba gear, compressed air guns, electric stun guns, driving a submarine using inclined fins, and undersea cities and farms (the last two have not, as yet, been confirmed as correct). As Brian Taves says in *The Jules Verne Encyclopedia*, Verne foretells "the development of motion pictures ... in *The Castle in the Carpathians* (1892)" ("Jules Verne: An Interpretation" 4) and an "enormous helicopter" in *Robur the Conqueror* (3).

Verne not only predicted the coming of the automobile, he also foresaw that it would, in his own words, "combat one of the greatest dangers of the future—the tendency of populations to desert the country and mass in great cities" ("Jules Verne's Autobiography" 55). As he went on to say, "The time is near at hand when every ordinary citizen may possess his auto—and with it will come an emigration back to the land. The rich will certainly take more and more to country life. For one thing they will feel more secure in the country—free from the reproachful gaze of the poor" (55).

However, Verne's most remarkable predictions were not concerned with land, sea, or air, but with space itself in *From the Earth to the Moon* (1865) and *All Around the Moon* (1870). In his biography of his grandfather, Jean Jules-Verne says, "Verne's eminently *scientific* [emphasis his] fiction never received a finer consecration.... 'It cannot be a mere matter of coincidence,' Frank Borman, the astronaut wrote to me in 1969. 'Our space vehicle [Apollo 8] was launched from Florida, like Barbicane's; it had the same weight and the same height, and it splashed down in the Pacific a mere two and a half miles from the point mentioned in the novel'" (93). "Verne's capsule" has the same "cylindro-conical shape" (Taves *The Jules Verne Encyclopedia* 2) as well. He correctly calculated escape velocity (12,000 yards per second) and a typical flight time (97 hours and 20 minutes) and suggested the use of rockets as retrojets, a suggestion picked up by Russian and German rocket scientists. Nor is this all. "Verne's astronauts experience weightlessness; they orbit at the same distance above the moon" (Taves 2). In addition, he saw "with uncanny accuracy the motivations for 'the space race' of the 1950s and 1960s" (Taves 2).

Jules Verne, himself, helped in the creation of this image, which eventually came to stand in the way of his purely literary ambitions and led him to say, "The great regret of my life is that I have never taken any place in French literature" (Sherard 115). For example, in an author's blurb for *The Adventures of Captain Hatteras* (1864), he wrote, "Mr. Jules Verne is the creator of a new genre.... The equal of our finest novelists, he is at the same time one of the best scientific minds of our times" (cited in Lottman 119). Most authors have to write such things, then as now, but the notion of Verne as scientist and his remarkable success in short- and long-term predictions were to cause him difficulties in his later career at a time when he was more interested in his reputation as an artist. Such perceptions also affected (and continue to affect) responses to Verne's work. This is at least a partial explanation for the attacks by critics who wish to elevate Verne's literary reputation on what they regard as the source of the problem.

There is indeed much to be said for examining Jules Verne as something other than a science fiction writer. The predictions he made and the scientific work he inspired aside, he spent most of his life foregrounding geography. Even the science fiction novels such as *Twenty Thousand Leagues Under the Sea* contain extended passages that anticipate Jacques Cousteau and *National Geographic*. In Arthur Evans's words, "Only about 20 of Verne's total literary output of 63 novels and 21 short stories ... could even be termed scientific in the technological sense. The bulk of his work resembles, rather, a gigantic travelogue—filled with visits to hundreds of foreign locales and offering lengthy and detailed descriptions of the geography, the indigenous flora and fauna, and the customs of the people living there" (*Jules Verne Rediscovered* 2). The destinations for all that travel are also significant, especially for a book like *The Castle in the Carpathians*, which might otherwise seem out of place in Verne's oeuvre. "The typical Vernian voyage may be defined in essence as a fictional journey to 'otherness'.... And among Verne's most favored settings in this regard are those that conform to the basic Romantic notions of exoticism: those that are far away (geographically and/or culturally), highly unusual, and picturesque.... Any milieu that exhibits a certain 'gothic' quality fits the description" (Evans *Jules Verne Rediscovered* 63).

The Castle of the Carpathians is a Gothic novel centered on a love story, and it has puzzled many people who expected something more like Verne's early novels or who came to it with the traditional picture of Jules Verne firmly in mind. Even the subject of love seems out of place. Charles Sheffield said about those early novels, "Verne's plots can be summarized as 'a bunch of cheerful but emotionally

challenged guys go off and have a rattling (and scientific) good time” (14). Again, Verne helped to imprint the image, “‘I’m very awkward about expressing feelings of love,’ he confessed to Hetzel [his editor and publisher]. ‘The very word “love” frightens me in the writing’” (Lottman 114).

Nevertheless, *The Castle of the Carpathians* is a Gothic novel, set in Transylvania, about two aristocrats obsessed with the same extraordinary woman. Verne also includes a secondary love affair between Nic Deck, a local forester, and Miriota Koltz, the daughter of the mayor. There is, as might be expected, a reasonable amount of Gothic (and Vernian) doubling. Each aristocrat is the last of his long line, each has a castle located in an inaccessible spot, and each shows symptoms of instability—or madness—from time to time. Brian Taves sees this doubling as contending. He writes in his review of the 1981 movie version, made in Czechoslovakia, “A man full of belief in the superiority of his era is lured into investigating curious phenomena, whereupon he is captured by a misanthropic, alienated outcast—of whom Captain Nemo was the prototype—and exposed to the scientist’s superior technology. The captive ... seeks to destroy this vision ... escaping to a world momentarily restored to a tentative status quo” (Taves “The Mysterious Castle in the Carpathians” 48).

While this version of the story is true, it is not the only version (or vision) of the narrative that Verne allows or even foregrounds. We are warned from the beginning that different people see the same events in different ways and that even the same people see the same events differently if their perspectives or technologies change. So the narrator of the novel says, “In the opinion of many tourists the Castle of the Carpathians existed only in the imagination of the country people” (Verne 49). And given the number of monstrously impossible things the country people imagine, that does not grant the title “character” much credibility. Frik, the local shepherd, sees the castle clearly (because of his good eyes) and believes in it and its legends, but even he cannot make out the smoke rising from it until he buys a cheap telescope from a passing peddler.

At the center of the story is La Stilla, an Italian singer of transcendent ability and beauty. She serves many purposes in the novel, not the least of which is her demonstration of the dangers that beset an artist who becomes famous, which may, in turn, make her a stand-in for Verne himself. La Stilla lives “only for her art,” and has never experienced love (Verne 150). She is, at the age of 25, an artist only and not a complete human being. Enter the villain and the hero of this novel or the two heroes or the two villains, depending on how one looks at them. Baron de Gortz, master of the Castle of the Carpathians, has been

watching her on the stage for six years, but he never approaches her or seeks to speak to her. He certainly seems to be the villain at first glance, "Every time La Stilla appeared, in no matter what theatre of Italy, there passed in among the audience a man of tall stature, wrapped in a long dark overcoat, and wearing a large hat which hid his face. This man would hurry to his seat in a private box previously engaged for him, and there he would remain, silent and motionless, throughout the performance" (Verne 151-152). Gradually, his nearly spectral presence works on La Stilla to such an extent that she contemplates retiring from the stage and giving up the only life she has.

Gortz is an open-ended menace. Is he a psychic vampire? He watches La Stilla with "his look ardently fixed on the artiste as if he would fascinate her" (172). Verne's "book," Brian Taves asserts, "revealed the combined influences of Jacques Offenbach, E. T. A. Hoffmann, Edgar Allan Poe, and the gothic" ("The Mysterious Castle" 48). In spite of his assertions about science, Verne opens up other, more fantastic possibilities, even in his descriptions of minor characters, "These vendors of thermometers, barometers, and cheap jewellery [sic] always seem to be a peculiar people and somewhat Hoffmanesque in their bearing. It is part of their trade. They sell time and weather in all forms—the time which flies, the weather which is, and the weather which will be—just as other packmen sell baskets and drapery" (41). La Stilla becomes weaker over time under Gortz's strange gaze. There is much talk of vampires in the novel; Frik claims he can control them (38). Is Baron Gortz the real thing? He clearly believes after her death that he is holding her soul in his hands.

Or does La Stilla die because the actual woman on whom she was perhaps based died young? Estelle Duchesne might have been Jules Verne's muse and/or mistress. "Her name would also throw light on the haunting figure of Stilla ... Estelle could be the model for the absent heroine of this novel, one of the most enigmatic and evocative in the Verne canon" (Lottman 117). In that case, Gortz becomes—literally—a stage prop.

But there is another man in this struggle for one woman. Count Franz de Télék is in some ways a younger, handsomer version of Baron Gortz. He is not, as Verne repeatedly calls Gortz, a "melomaniac," but his passion for La Stilla is equally monomaniacal. He, however, wishes to possess her and not merely her voice, and when she chooses to abandon her career and marry him, she commits herself to living in a castle in "one of the chief towns of the State of Roumania, which borders the Transylvanian provinces south of the Carpathian chain" (Verne 133). In her desperate desire to escape what she perceives as the baleful influence of Baron Gortz, La Stilla chooses Franz because he promises her

love and safety, not because of any love she feels for him or because she wishes to abandon her career.

She has become, paradoxically, the object of the voyages of discovery that the Count and the Baron make from their exotic castles to the more mundane location of Italy. As Roland Barthes argues in "The *Nautilus* and the Drunken Boat," Verne's "work proclaims that nothing can escape man, that the world, even its most distant part, is like an object in his hand, and that, all told, property is but a dialectical moment in the enslavement of Nature" (65). La Stilla, is, we are told, an incredible example of Nature equaling art: she "was a woman of ideal beauty, with her long golden hair, the ardour of her deep-black eyes, the purity of her complexion, and a figure which the chisel of a Praxiteles could not have made more perfect" (Verne 149-150). In this vision of the story, La Stilla, as property, is meant to be taken back to a safe place—a castle—as wife by Franz or musical soul by Gortz. In either case she would come to occupy "an absolutely finite space" (Barthes 66), which confers the safety and seclusion she is promised but precludes the possibilities and even perhaps the personality she currently possesses. In the event, the Castle of the Carpathians fails as a protective structure. It may, even more than the *Nautilus*, deserve the description "the most desirable of all caves" (66), but as so often happens in Verne, the struggles of men destroy the best achievements of mankind.

If Franz were, in fact, the hero of the novel, would he not treat her as something other than an acquisition, something more than a prize to be carried off? Would he not make some attempt to save her life as she knows and lives it? In other words, would he not try to win her love by saving her musical life? It is the only thing she has ever cared about, the only happiness she has ever had, and as far as we can tell, the only ability she displays. With all his youth and money and passion, could he not somehow challenge Gortz and forbid him from haunting La Stilla's performances? Of course, if Franz were to solve her problem, he could not take advantage of it. Previously, La Stilla has not only denied him affection, she has forbidden him access, despite his attempts. She is caught between two appropriately Gothic aristocrats, and in the end she makes a desperate choice. Can either one of them be labeled a hero? Are they both villains or are they merely caught in the inevitable, intractable suffering that attaches to humanness?

The Gothic novel represented a sort of breakdown in Enlightenment confidence in the rationality of man and the inevitability of progress and, consequently, a breakout for those dark elements of the psyche that the French *philosophes* had tried to leave behind—literally in the past and imaginatively in their visions of what the present and

future would be like. As Derek Offord writes, “Gothic fiction ... gropes for a formulation of the anxieties to which the French Revolution—among other factors, such as industrial revolution, urbanisation and attendant economic and social change—had given rise” (52-53). By creating a Gothic narrative, Verne had opened up that vocabulary of criticism and its topics, including mining, pollution, the vagaries of international finance, and the invincible ignorance of an impoverished population.

Verne tells us simultaneously that, “We live in times when everything can happen—we might almost say everything has happened. If our story does not seem to be true to-day, it may seem so to-morrow, thanks to the resources of science, which are the wealth of the future” (*The Castle of the Carpathians* 33), and “at the same time it is as well to note that Transylvania is still much attached to the superstitions of the early ages” (33). In fact, “Civilization is like air or water. Wherever there is a passage, be it only a fissure, it will penetrate and modify the conditions of a country. But it must be admitted that no fissure has yet been found through this southern portion of the Carpathians” (64).

Verne uses this contrast to brilliant effect in—as might be expected—a variety of ways. The little village of Werst, which is just within eyeshot of the Castle of the Carpathians, is filled with peasants who believe in every monster about whom anyone ever told a tale. They are so superstitious that they try not to go out on Thursdays—since that is an especially active time for the demonic and undead. When it becomes clear that someone needs to investigate what’s happening in the Castle, the villagers provide some of the funniest cowards since Falstaff, plus the closest thing to a hero that the book can truly offer.

The villagers of Werst see the Castle of the Carpathians as terrifying, evil, literally filled with demons.

“‘Yes, my good friends,’ said the innkeeper. ‘If there is a smoke from the donjon chimney, it is because there is a fire, and if there is a fire it must have been lighted by a hand—’

‘A hand!—at least a claw!’ said an old peasant, shaking his head” (Verne 77).

The final confrontations take place in the labyrinthine passages of a structure that Verne deliberately compares to the original labyrinth built by Daedalus (182). At the heart of the mystery of this Gothic novel is electricity, a light that in Verne’s other stories (and here too) can do anything. And by that light Verne reveals that perhaps the ultimate

explanation for what happens, the evil that the villagers fear, and the ineffable terror that drives real horror novels is to be found, not in vampires or other demons, but in the two human beings fighting alone at night over what's left of the woman they have killed between them. One of them goes mad and kills himself because he has lost the last sounds she made in life; the other goes mad but is brought back to sanity by the songs he wanted her to give up singing. What inhuman monster could do more?

And yet, Jules Verne may have a still more complex message. Two men are driven mad by the voice of an authentic siren. There are many such creatures in Slavic legends. The local schoolmaster, Magister Hermod, who is an expert on legends and penmanship and nothing else, warns early in the story that such things may be in the Castle, "Why should they not be spirits, goblins, perhaps even those dangerous lamias which present themselves under the form of beautiful women?" (78). La Stilla inspires love but does not feel it. She has the irresistible power of art and the glamour of extreme celebrity. Such a combination could drive people mad in real life as well as in fiction, as Verne well knew.

Franz de Télék feels "as if some invisible bond he could not break had attached him to the singer" (150). After her death, he nearly dies as well. When he recovers, he insists on visiting her grave, "There Franz threw himself on the cruel ground—he would have torn it up with his finger-nails to bury himself by her side" (160). Gortz uses the sound of her voice and her image to lure Franz to the Castle. He, himself, is arguably trapped there by the recordings of that magical voice. Here is a magic inside a mystery—something seemingly alive that is actually dead, and, indeed, it is worth noting in this context that not all early reactions to the phonograph were positive. As John M. Picker writes in "The Victorian Aura of the Recorded Voice," "Endless repetition of a disembodied voice had the potential to distort even the most benign speech into a monotonous rant that sounded diabolical, perhaps even terrifying" (769). However it may have sounded to Gortz, he was repeating La Stilla's last song over and over. It is a siren song that reaches beyond the grave. In Verne's last vision of him, "He uttered a terrible cry.

'Her voice—her voice!' he repeated. 'Her soul—La Stilla's soul—it is ruined—ruined—ruined!'

And then with his hair bristling and his hands clenched, he was seen to run along the terrace, shouting,—

'Her voice—her voice! They have taken away from me her voice! Curse them!'" (229-230).

Orfanik, that scientist without conscience and servant without loyalty to his master the Baron, would seem to see La Stilla as more monster than muse, though both can be perilous, "Orfanik stated that La Stilla was dead, really dead and—such was his expression—buried, and well buried, for more than five years in the cemetery of Santo Nuovo Campo at Naples" (233). And buried again in the ruins of the Castle of the Carpathians, although her voice and its power "live" on beyond the story in that other castle where Franz may still be suffering from "the delirium of his idolatry for La Stilla" (151), still caught by the disembodied sound of the body he hoped to possess. Barthes says, "Verne in no way sought to enlarge the world by romantic ways of escape or mystical plans to reach the infinite: he constantly sought to shrink it, to populate it, to reduce it to a known and enclosed space, where man could subsequently live in comfort" (65-66). But, as often happens, Verne is wiser than his interrogators. In the little rooms to which Verne's heroes and antiheroes retire, it is not comfort they find.

I have one last note about the power and complexity of *The Castle of the Carpathians*. I began this long essay by discussing Jules Verne as a battleground for critics. There may not be a real original for La Stilla, but what if there is and what if that original was not Estelle but Jules Verne himself? La Stilla is caught between fanatic admirers. Each of them seeks to possess her in a different way while her goal is only to produce more of her music and not to interact with her audience in any real sense at all. During her lifetime and after her death, La Stilla's fans fought to claim her and even to determine what she should be. At what point does interpretation become appropriation? Jules Verne certainly knew what it was like to face unreasoning and uncontrollable adulation. He knew how dangerous such passions could be. After all, in 1886, he had been shot by a deranged nephew, and it was a wound from which he never wholly recovered. Gaston, his brother Paul's child, had supposedly said, "I wanted to kill uncle Jules because he is so good that he must go to Paradise right away" (Lottman 257). Jean Jules-Verne suggests that "Gaston *may* have felt smothered by his uncle... because he was famous" (160). The pain of the injury would have been with Verne as he wrote this novel. Is *The Castle of the Carpathians* one more extrapolation from a painful present reality into what he thought the future might be? If so, if he foresaw fanatical admirers struggling over what was left of his voice, he could claim one more accurate prediction.

Works Cited

Angenot, Marc. "Jules Verne and French Literary Criticism." *Science Fiction Studies* 1.1 (Spring 1975): 33-37.

Asimov, Isaac. *Asimov on Science Fiction*. New York: Doubleday & Company, Inc., 1981.

Barthes, Roland. "The *Nautilus* and the Drunken Boat." *Mythologies*. Trans. Anette Lavers. New York: Hill and Wang, 1972. 65-67.

Beschloss, Michael R. *The Crisis Years: Kennedy and Khrushchev 1960-1963*. New York: Edward Burlingame Books, 1991.

Evans, Arthur B. "A Bibliography of Jules Verne's English Translations." *Science Fiction Studies* 32.1 (2005): 105-141.

Evans, Arthur B. *Jules Verne Rediscovered: Didacticism and the Scientific Novel*. New York: Greenwood Press, 1988.

Jules-Verne, Jean. *Jules Verne: A Biography*. New York: Taplinger Publishing Company, 1976.

Lottman, Herbert R. *Jules Verne: An Exploratory Biography*. New York: St. Martin's Press, 1996.

Miller, Walter James, and Frederick Paul Walter, transl. Introduction. *Jules Verne's 20,000 Leagues under the Sea: The Completely Restored and Annotated Edition*. Annapolis, Maryland: Naval Institute Press, 1993.

Offord, Derek. "Karamzin's Gothic Tale: *The Island of Bornholm*." *The Gothic-Fantastic in Nineteenth-Century Russian Literature*. Ed. Neil Cornwell. Atlanta, GA: Rodopi, 1999. 37-59.

Okuda, Michael, and Denise Okuda. *The Star Trek Encyclopedia: A Reference Guide to the Future*. New York: Pocket Books, 1999.

Picker, John M. "Aura of the Recorded Voice." *New Literary History* 32.3 (2001): 769-786.

Sheffield, Charles. *Borderlands of Science: How to Think Like a Scientist and Write Science Fiction*. New York: Baen, 1999.

Sherard, R.H. "Jules Verne at Home: His Own Account of His Life and Work." *McClure's Magazine* 2.2 (January 1894): 115-124.

Taves, Brian, and Stephen Michaluk, eds. *The Jules Verne Encyclopedia*. London: The Scarecrow Press, 1996.

Taves, Brian. "The Mysterious Castle in the Carpathians by Oldrich Lipský: Jan Svankmajer." *Film Quarterly* 38.2 (Winter 1984-1985): 48-50.

Unwin, Timothy. "Jules Verne: Negotiating Change in the Nineteenth Century." *Science Fiction Studies* 32.1 (2005): 5-17.

Verne, Jules. *The Castle of the Carpathians*. Forest Tsar Press, www.foresttsarpress.com, 2010.

Weber, Eugen. Introduction. *Paris in the Twentieth Century*. By Jules Verne. New York: Random House, 1996.

Nostalgic Renderings of Victorian Clothing in John Fowles's *The French Lieutenant's Woman*, A.S. Byatt's *Possession*, and Neal Stephenson's *The Diamond Age*

Amanda Scott

Dixie State College

Abstract

The Neo-Victorian Fiction of Fowles, Byatt, and Stephenson relies heavily on the Victorian visual aesthetic in portraying characters' clothing. These representations, however, are tainted by modern stereotypes of the Victorians and manifest the socio-cultural distinction the authors draw between the Victorians and modern civilization.

The Neo-Victorian novel, which emerged in the latter half of the 20th century, is a dynamic genre that rewrites the past and compares the postmodern era with the Victorian era. Neo-Victorian novels generally emphasize the tensions between postmodern values and Victorian values. Texts often represent this tension when narrators inform their perceptions of the Victorian Era with postmodern philosophies or draw on later history to contextualize Victorian events. Although Neo-Victorian

novels generally emphasize moral differences, aesthetic differences also prove key in contrasting the postmodern and Victorian eras. In John Fowles's *The French Lieutenant's Woman*, A.S. Byatt's *Possession*, and Neal Stephenson's *The Diamond Age*, clothing functions as an important conduit for both Victorian values and aesthetics. These novels often convey the rigidity of Victorian values through the modest and restrictive fashions of the era. Victorian fashions are not conveyed as necessarily negative though; rather, the Neo-Victorian novels examined here exhibit nostalgia for Victorian values through clothing.

Throughout the past decade, many critics have explored the treatment of Victorian values in the above texts,¹ but few have mentioned clothing, and none have investigated the connection between Victorian aesthetics and Victorian values within these texts in depth. Critics concur on one point though; the Neo-Victorian novel draws attention to how postmodern preoccupations taint Neo-Victorian representations of the Victorian Era and its corresponding values. Analyzing the interest of authors of Neo-Victorian texts in the Victorian, Jess Nevins states, "The attraction of the surface elements of the Victorians, the trappings and visual style, is obvious, but the 19th century has a further appeal for post-modern writers." Nevins goes on to emphasize that the "surface elements" of the Victorian era and the moral values of the age are utilized in Neo-Victorian texts for separate purposes (8). This essay, in contrast, argues that Victorian "surface elements" and Victorian values are crucially connected in Neo-Victorian novels. Victorian clothing proves key precisely because of the values it embodies.

The relationship between the Victorian and the postmodern is complicated and reflexive. The Neo-Victorian novel not only draws on the similarities between the two eras, as Nevins suggests, but also conceptualizes Victorian aesthetics and values subjectively, through postmodern ontologies. Concerning this phenomenon, Sarah Gamble writes, "[W]hen we look at the Victorians, we see a world whose sexual and social mores are both quaint and titillating ... The function of the neo-Victorian novel may be to animate the past, but it can only do so from the perspective of the present, which will always read it as reflective of its own preoccupations" (127). Thus, postmodern representations and general understanding of the Victorian era is tainted by the postmodern experience; we cannot remove ourselves entirely from postmodern ontologies to understand the Victorian era.

¹ For examples of strong criticism on the aforementioned texts, see Peter Brigg's "The Future as the Past Viewed from the Present: Neal Stephenson's *The Diamond Age*," Daniel Punday's "Meaning in Postmodern Worlds: The Case of *The French Lieutenant's Woman*," and Dana Shiller's "The Redemptive past in the Neo-Victorian Novel."

The intersection between Victorian aesthetics and postmodern philosophies reflects in many mainstream and counter-cultural fashion trends, which have been popularized in the past 15 years. During the 1990s, Victorian aesthetics re-emerged in high fashion and Neo-Victorian literature peaked in popular fiction. With the resurgence of Victorian aesthetics in the fashion world, counter-cultures have formed in reaction to the Victorian aesthetics, such as Victorian Gothic and Steampunk stylings. These counter-cultural phenomena reinvent conservative fashions into edgier modern ones. While Neo-Victorian and Victorian Gothic aesthetics largely reproduce Victorian trends with modern flourishes, Steampunk aesthetics specifically use modern technological devices to reinvent items with traditional Victorian styling.

It is no surprise, then, that Victorian fashions are so central to representing Victorian aesthetics in the texts of Fowles, Byatt, and Stephenson. These texts undeniably represent the Victorian era with preoccupations and contain the concept of the period through postmodern philosophies. Namely, what the texts and modern-day fashions label as “Victorian aesthetics” are exaggerated stereotypes. A similar phenomenon has occurred in conceptualizing Victorian moral values. Victorian values were, no doubt, much more rigid than the relative moralism common in postmodern philosophies; however, upon closer study of the Victorian era, one will find that society was much more sexually explorative than the stereotype suggests.

All three texts exhibit desire for this postmodern construction of “Victorian values” through clothing, and while each novel represents Victorian clothing differently, each text exhibits desire for the clear boundaries of morality. Neo-Victorian novels, however, present modest Victorian clothing not as a helpful protection to enforcing morality, but as a titillating boundary to transgress. Understandably, the transgressed boundary usually results in a sexual encounter, but this transgression of Victorian barriers also relates to working through the barrier of Victorian conservatism to reach a new political or cultural order.

The French Lieutenant's Woman, published in 1969, debuted as one of the first popular Neo-Victorian novels. The narrative tracks the scandalous affair between Charles and Sarah—a relationship that ends Charles's engagement with the well-off Ernestina and brings social shame on them both. At first glance, Fowles's novel does not seem to emit nostalgia about the Victorian Era at all, but seems preoccupied with stereotyping and mocking Victorian values (Goscilo 65). The narrator also seems to mock the Victorian era as it presents postmodernism, and particularly existentialism, as superior to the Victorian (Gauthier 37). Fowles presents the Victorians as philosophically and scientifically deprived; one of their only good traits seems to be the

hypersexuality that occurs when a rebel comes out of suppression and chooses to transgress the era's concrete moral boundaries. By the text presenting moral boundaries as erotic, however, Fowles's novel does prove nostalgic for Victorian values, in a sense.

Clothing is particularly important for the text's representation of Victorian values. Clothing often functions to mark aesthetic trends, but there is always a deeper cultural implication from these aesthetics. In the first few pages of the novel, Victorian clothing serves to establish the year of the fictional present, 1867. The narrator mentions the technology of aniline dyes apparent in the characters' bright clothing and emphasizes the move away from crinoline and large bonnets in his setting (5). The Victorian aesthetic proves crucial in establishing the period's cultural milieu. The narrator encapsulates and reduces the fictional present, in part, through Victorian fashion. The clothing helps to stereotype and continually emphasize the naïve contained identity of his unphilosophical Victorians.

Another manifestation of the narrator's cynicism toward Victorian values appears in the easy categorization of the novel's characters. Clothing serves as a primary marker of characters' self-assured identity and position in society. For example, the text defines the outcast, Sarah, through her distinctly un-Victorian, poorly fitting clothes. The narrator frequently mentions Sarah's clothes to stress how out of place Sarah is in her environment: "[Charles] perceived that the coat was a little too large for her, and that the heels of her shoes were mudstained" (86). The text seems to rely on Sarah's clothing and appearance to define an otherwise enigmatic woman. Charles is able to make a value judgment of her as a misfit, fallen woman through her choice of attire. Later, the strangeness of her clothes is again emphasized: "That strange-cut coat, that black bonnet, that indigo dress with its small white collar" (275-76). The narrator continually defines Sarah as unordinary and unfitting in her Victorian world; her clothing represents her social standing and reveals aspects of her personality. She is independent and unconcerned with adhering to any sort of contemporary aesthetic.

The text especially utilizes clothing to define and explain Sarah's character toward the end of the novel when she is presented as an emancipated New Woman. The narrator posits,

This was someone in the full uniform of the New Woman, flagrantly rejecting all formal contemporary notions of female fashion. Her skirt was of a rich dark blue and held at the waist by a crimson belt with a gilt star clasp; which also enclosed the pink and white striped silk blouse, long-sleeved, flowing, with a delicate small collar of white lace, to which a small

cameo acted as tie. The hair was bound loosely back by a red ribbon. (431)

This description of Sarah serves to both date the narrative and to contain Sarah's social position. Sarah's colorful and less restrictive clothing, her "striped silk blouse" and "flowing" skirt, indicate a Pre-Raphaelite aesthetic as was represented in fine art toward the end of the Victorian era. This colorful fashion aesthetic was associated with sensuality and free love in newly liberated New Women.

This passage also illustrates how clothing serves to contain and stereotype Sarah. Simply by rejecting contemporary fashion and stylistically resembling a New Woman, the narrator easily categorizes her as an emancipated woman. The text, however, never allows Sarah a clear voice to declare her emancipation and insistently reduces and marginalizes her. She is, essentially, a flat character. As Goscilo posits, "inter-textual tensions...betray [Fowles's] own limitations in depicting womanhood" (70). Although Sarah may seem like an unconventional and forward-thinking member of society, she actually serves as the typical fallen New Woman character. Descriptions of Sarah's clothing aid the narrator in stereotyping Sarah.

The author's use of clothing as a means of defining character types and containing the aesthetic values of the period hardly makes the novel nostalgic for Victorian values. On the contrary, the novel presents its stereotyped characters as cogs in the moral Victorian machine. The novel utilizes clothing in another way, however, which is incredibly nostalgic: as an erotic barrier.

When Sarah and Charles consummate their relationship, the turning point of the scene is when Sarah's green shawl falls from her shoulders, then "...[A] whole ungovernable torrent of things banned, romance, adventure, sin, madness, animality, all these things coursed wildly through him." (349). As the sexual encounter ensues, Charles's clothing is a barricade that at once frustrates and thereby enhances sexual pleasure and urgency (349). As Charles removes his clothing, each Victorian item is listed—the frock coat, waistcoat, pearl tie pin, and cravat. The narrator emphasizes the multilayered clothing to suggest the stifling barrier it presents to sexual expression. It is clear that as Charles strips down these layers of clothing (and suppression) he is also stripping back down to primitive sexual instincts.

In this moment of the text, a clear tension emerges between the Victorian values that forbid Charles and Sarah's sexual encounter and the voyeuristic postmodern narrator who seems to take pointed delight in the transgression of moral boundaries. The novel, however, does not seem to posit that the transgression of Victorian mores is necessarily

existentialism prevailing over bigoted moral standards; rather, the narrator takes delight in the boundary. The passion of the encounter is enhanced by the moral boundary, which the postmodern narrator seems to relish. So, while the text disparages Victorian values, overall, it reveals nostalgia for the sexual boundaries of the time. The boundaries are commendable, not because the postmodern narrator privileges them morally, but because they enhance sexual encounters. Thus, while Fowles's novel contains a clear criticism of Victorian values in seeing them as restrictive and unnecessary, the novel also exhibits clear nostalgia for forbidden sexual passion.

A.S. Byatt's *Possession* uses clothing in a similar way to *The French Lieutenant's Woman*. Clothing often serves to date the parts of the narrative set in the 19th century and also serves as an enticing sexual barrier. The novel's attitude about Victorian values, however, is drastically different from the negative containment and stereotyping of the era in Fowles's novel. In the text, Maud and Roland, two 20th-century scholars, become consumed with desire to understand two Victorian poets, R.H. Ash and Christabel LaMotte, and their forbidden romance. One reason Ash and LaMotte's story fascinates Maud and Roland so much is because of the stable sense of self the Victorian characters maintain. As Martin Löschnigg posits, "descent into the past becomes," for Maud and Roland, "a descent into self, bringing to light ... a desire for identity" (qtd. in Steveker 124). *Possession* emits a strong nostalgia for Victorian values and what the novel's 20th-century characters deem a stable sense of identity.

Clothing very much contributes to the novel's nostalgia. Where Fowles's novel uses Victorian clothing as a means of stereotyping characters, Byatt's novel refrains from this sort of simplification. Instead, clothing functions to give vivacity and a romanticized beauty in describing the Victorian Era. In the novel's first passage set in the 19th century, the narrator describes the vibrant detail of Christabel's attire:

she wore a grey-striped muslin dress over which she had cast an Indian shawl with marine-blue and peacock paisleys on a dove-grey ground; she had a small grey silk bonnet, under the brim of which appeared a few white silk rosebuds... Everything about her was both neat and tastefully chosen, breathing no hint of extravagance, but betraying no signs of poverty or skimping to the curious eye. Her white kid gloves were supple and showed no signs of wear. Her little feet, which appeared from time to time as the carriage movement displaced the large bell of her skirt, were encased in a gleaming pair of laced boots in emerald green leather. (299)

This passage strongly illustrates Victorian aesthetics. Each detail here, “the grey-striped muslin dress,” the exotic-looking “Indian shawl,” the “white kid gloves,” “the large bell of her skirt,” and her “gleaming pair of laced boots,” all represent vibrant Victorian fashions. Unlike the descriptions in Fowles’s novel, however, Christabel’s aesthetic choices do not necessarily categorize her. The narrator describes her clothing as “neat and tastefully chosen ... but betraying no signs of poverty.” While the narrator makes aesthetic judgments about Christabel, she is certainly not stereotyped by her choice of clothing. Clothing here serves to date the period in a far less reductive way.

Also similar to Fowles’s novel, clothing serves as an enticing barrier to sexual expression. When Randolph (R.H.) and Christabel consummate their relationship, clothing not only marks the era, but represents desire and eroticism. After Christabel has undressed, Byatt presents the image of her discarded clothes: “On one chair stood a kind of trembling collapsed cage, the crinoline, with its steel hoops and straps. Under it, the small green boots” (308). The narrator emphasizes the restrictive nature of Christabel’s cage crinoline here. The “trembling collapsed cage,” so clumsy and annoying to many Victorian men, as Steele suggests, is now shed to enable intimacy (79). Although stripping away clothing likely occurs in most sexual encounters, here the particular restrictive Victorian crinoline is exceptionally hindering to sexual passion, and thus, the shedding of it is particularly erotic.

In another moment in the text, this time set in the 20th century, Roland is interested in having a sexual encounter with his girlfriend, Val, who lies naked in bed beside him. At first he is unsure whether the encounter will “work,” but then he remembers a sketch of the Victorian woman, Ellen Ash, that he saw in a book. The narrator describes her as, “a woman not naked but voluminously clothed, concealed in rustling silk and petticoats, finger folded over the place where the tight black silk bodice met the springing skirts ... a stiff bonnet framing loops of thick hair ... The mental vision of this woman, half-fantasy, half-photogravure, was efficacious” (141). Clearly, Roland is excited by the very modesty of Victorian garb. The physical barrier of “rustling silk and petticoats” and the “tight bodice and springing skirts” seems much more arousing than his girlfriend’s nakedness. Similar to Charles and Sarah’s sexual encounter in *The French Lieutenant’s Woman*, restrictive Victorian clothing here serves as a physical barrier. The fantasy and excitement that Victorian clothing creates for Roland represents a postmodern nostalgia for more concrete sexual mores, even if the sole desire to have the mores is the thrill of transgressing them.

The nostalgia for Victorian values and what characters perceive as stable identity looms large in *Possession*. Clothing serves as one device

that illuminates 19th-century England and its contrast to the postmodern world. The vivid aesthetic detail of the period gives the novel's characters surface detail, but does not define characters' societal position. All in all, Byatt's text proves much more clearly nostalgic toward Victorian values than does Fowles's novel. The text functions similarly, however, in that modest Victorian garb is titillating from the postmodern perspective, whether demonstrated through Fowles's existentialist narrator or Byatt's postmodern intellectual scholar. Both texts emit a particular nostalgia in regards to the value of boundaries and the pleasure of transgressing them.

Neal Stephenson's *The Diamond Age*, like the two aforementioned novels, clearly portrays nostalgia for Victorian values, but Stephenson's delivery is vastly different. While *The French Lieutenant's Woman* is set in the 19th century and *Possession* is set in both the 19th and 20th centuries, *The Diamond Age* presents a futuristic Neo-Victorian society in the late twenty-first century. Stephenson's text is a Steampunk science fiction novel that conceptualizes the future world as divided into phyles (tribe-like formations). One such phyle is the Neo-Victorians—people who act and dress like Victorians, but possess much more advanced technology.

The plot revolves around Nell, a girl who rises from an abusive, impoverished family to a Neo-Victorian way of life. A magical, interactive primer academically and morally educates Nell and prepares her for life with the Neo-Victorians. The Neo-Victorian lifestyle is not the end of Nell's journey, however; she eventually rises from Neo-Victorian conservatism to what the text presents as a new, higher order.

The Diamond Age proves, no doubt, the most nostalgic toward Victorian values of any of the three novels examined in this essay. The reasons for nostalgia are entirely different though. While the first two novels are preoccupied with the sexual tension aroused by Victorian moral boundaries, Stephenson's text largely implies the worth of Victorian values for their political effects. This text posits Victorian conservatism as a gateway to a higher order that escapes traditional morality (Berends 18).

Even so, the novel does not only uphold Victorian values, but also upholds Victorian aesthetics. The novel portrays a Steampunk, science-fiction aesthetic that involves repurposing traditional Victorian items with postmodern technological advancement, while still maintaining the items' original aesthetics. This Steampunk repurposing creates a

*bricolage*² of the old and the new and endows the novel with a warped concept of time.

One Victorian item repurposed as Steampunk in the text is the pocket watch. This item seems to be fetishized, as it is not desired for its use value but for its antique qualities. In one instance, the Neo-Victorian Equity Lords are described as “[those] executives, whose gold watch chains, adangle with tiny email-boxes, phones, torches, snuffboxes, and other fetishes, curved round the dark waistcoats they wore to deemphasize their bellies” (10). The Neo-Victorian men carry pocket watches, and while pocket watches are not exclusively Victorian, they are strongly linked to Victorian aesthetics for this phyle. The watches seem fetishized for their antiquity, but the futuristic world adds various functional technologies to the watches, like “tiny email-boxes, phones, torches, [and] snuffboxes.” Postmodern technologies impose on an antiquated item. We see the pocket watch produced in a fetishized way again when Hackworth is getting ready for the day. “He was stringing his watch chain around various tiny buttons and pockets in his waistcoat. In addition to the watch, various other charms dangled from it, such as a snuffbox that helped perk him up now and then, and a golden pen that made a little chime whenever he received mail” (31). The Steampunk watch expresses nostalgia for Victorian aesthetics, but the watches are also endowed with postmodern conveniences.

Stephenson reformulates many other Victorian clothing items to have postmodern technological properties. The Neo-Victorians wear white kid gloves “constructed of infinitesimal fabricules that knew how to eject dirt” (9). This reformulation again involves a Victorian aesthetic with technological improvements. Another reformulation is women’s veils, which can shield the wearer with a “field of microscopic, umbrellalike aerostats programmed to fly in a sheet formation.” The narrator indicates that the veils offer women “protection from unwanted scrutiny” (300). This reformulation suggests nostalgia for modesty. The Neo-Victorian women do not see the veil as restrictive, but as liberating because of the privacy it provides. Furthermore, it is the technological advancement paired with advanced technology that provides particular power and protection.

Another important reformulation in the novel is that of the Victorian corset. In one passage, the narrator describes Gwen’s tight-bodied

² *Bricolage* is a term popularized in literary studies by Jean-Francois Lyotard. This term has multiple meanings in various post-modern studies. It comes from the French verb *bricoler*, the contemporary meaning of which is “to do yourself.” Lyotard theorized that in post-modernity we combine aesthetics and cultural properties from multiple eras of history into one post-modern aesthetic.

Victorian dress as being sexual and revealing, rather than modest and chaste, as it is traditionally considered. The traditional corset that enabled the tight bodice was considered necessary for middle and upper-class women to be properly contained and covered (Miller 129-30). Leslie Shannon Miller states, "The iron cage of the corset creates a 'touch-me-not' aura of high morality and discipline that went hand in hand with the message of physical youth indicated by the wearer's slim waist" (137). Gwen's corseted attire, in contrast, consists of "modern fabrics thinner than soap bubbles," and, rather than concealing the female body, "made everything obvious." Also, rather than the traditional idea that the corset enabled a woman to have the ideal figure, Gwen's corset is not restraining, but only revealing (Stephenson 29-30).

This Steampunk rendering of the Victorian corset suggests a subversion of traditional modesty. Although the Neo-Victorian women still value the aesthetic of the corset, they desire to wear it without restriction and in a way that reveals their bodies. Here the text presents a repurposing of both a Victorian clothing item and corresponding value into a sexual aesthetic, more conducive to the futuristic Neo-Victorian world.

Stephenson's Steampunk reformulation of Victorian aesthetics, then, emphasizes the liminal trisection of past, the author's present, and the imagined future. In true science-fiction formulation, Stephenson takes old objects and makes them edgy and new in a way that combines the familiar and the foreign for his audience. The reformulated Victorian items, in turn, never function as mere surface elements, but correspond with Victorian values. By repurposing Victorian aesthetics through futuristic technology, the novel suggests a repurposing of outdated Victorian values with postmodern infusions. The combination of the conservative and the postmodern seems to be the privileged order here—both aesthetically and philosophically.

The Diamond Age, however, is most nostalgic toward Victorian values through its political rendering of the Neo-Victorians. The novel tracks Nell's ascension from a depraved childhood environment to an education through the magical Primer book, which teaches Nell Victorian morals and trains her to be gentle and cultured. The text, then, posits Victorian values as the most meaningful ideological milieu for Nell to transcend. The new order involves Nell rising up to lead a female Chinese army.

At the conclusion of the novel, Nell descends into the ocean to rescue her pseudo parents, Miranda and Carl, in a moment of victory and a show of female strength. At this liberating moment, she has shed her Neo-Victorian clothing: "Princess Nell walked naked into the sea at dawn, vanished beneath waves turned pink by the sunrise, and did not

return" (450). Although she is naked and vulnerable, by Victorian standards, the novel portrays her as a warrior and leader who is set free by her nudity. It is clearly through escaping clothing, and particularly Neo-Victorian clothing, that Nell can exercise her power as the leader of a new order. Quite unlike the case in Fowles's and Byatt's novels, the removal of Victorian clothes is not hyper-sexualized. *The Diamond Age* does not present Nell as a sex object in her nude plunge into the sea, but as a powerful woman. The treatment of characters' shedding Victorian clothing is similar in all three novels, however, in that the nudity allows for more free expression and distance from constructed social values. Furthermore, once Nell sheds her Neo-Victorian clothing she is able to take on her new quest. The presence of the value system and the clothing seems crucially linked.

Nell's escape from Victorian values does not necessarily discount Neo-Victorian conservatism, however. It is through the contained mores of Victorian ontology, and not the chaos of the futuristic world, that Nell can transition to true heroism (McClancy 73-76). Thus, *The Diamond Age* emotes nostalgia for both Victorian aesthetics and values, even if it is only valuable to the novel's protagonist as a transitional ideological force.

To conclude, each of the novels analyzed above emotes nostalgia for Victorian values. *The French Lieutenant's Woman* exhibits ambivalence toward Victorian values, clearly disparaging the philosophically limited lives of its Victorian characters and easily stereotyping these characters. In contrast, Byatt emotes reverence for Victorian values and aesthetics and is much more overtly nostalgic. Both of these novels portray titillation at the concept of Victorian values and the possibility of transgressing them. Clothing also serves as a strong indicator in both describing and defining the Victorian era and representing set sexual values. *The Diamond Age* also proves nostalgic in portraying Victorian values as a transitional means to a higher order. Tellingly, nostalgia toward Victorian values never stands as final trajectory point in these novels; rather, Victorian values serve as a threshold for either sexual passion or political enlightenment. Clothing, then, works as strong indicator of Victorian values in these novels. Only through stripping away the layers of Victorian clothing can the characters in these novels transgress or overcome rigid Victorian mores.

Works Cited

Berends, Jan Berrien. "The Politics of Neal Stephenson's *The Diamond Age*." *New York Review of Science Fiction* 9.8 (1997): 15-18.

Literature Resource Center. Web. Accessed October 10, 2010.

Byatt, A. S. *Possession: a Romance*. New York: Vintage, 1991. Print.

Fowles, John. *The French Lieutenant's Woman*. Boston: Back Bay, 1998. Print.

Gamble, Sarah. "'You Cannot Impersonate What You Are': Questions of Authenticity in the Neo-Victorian Novel." *Lit: Literature Interpretation Theory* 20.1 (2009): 126-40. *LION*. Web. Accessed October 13, 2010.

Gauthier, Tim S. *Narrative Desire and Historical Reparations: A.S. Byatt, Ian McEwan, Salmon Rushdie*. New York: Routledge, 2006. Print.

Goscilo, Margaret Bozenna. "John Fowles's Pre-Raphaelite Woman: Interart Strategies and Gender Politics." *Mosaic* 26.2 (1993): 63-82. *Literature Resource Center*. Web. Accessed October 20, 2010.

McClancy, Kathleen. "Reclaiming the Subversive: Victorian Morality in Neal Stephenson's *The Diamond Age*." *Tomorrow through the Past: Neal Stephenson and the Project of Global Modernization*. Ed. Jonathan P. Lewis. Newcastle-upon-Tyne: Cambridge Scholars, 2006: 69-85. Print.

Miller, Leslie Shannon. "The Many Figures of Eve: Styles of Womanhood Embodied in a Late-Nineteenth-Century Corset." *American Artifacts: Essays in Material Culture*. Ed. Jules David Prown and Kenneth Haltman. East Lansing: Michigan State U Press, 2000: 129-48. Print.

Nevins, Jess. "Introduction: The 19th Century Roots of Steampunk." *Steampunk*. Ed. Ann VanderMeer and Jeff VanderMeer. San Francisco: Tachyon Publications, 2008. Print.

Stephenson, Neal. *The Diamond Age*. New York: Bantam, 2000. Print.

Steveker, Lena. *Identity and Cultural Memory in the Fiction of A.S. Byatt: Knitting the Net of Culture*. Basingstoke, England: Palgrave Macmillan, 2009. Print.

Instantaneous Frequency Tracking of Partial of a String Vibrating at Large Amplitudes Using a High-speed Camera

David C. Ripplinger, Brian E. Anderson, and Timothy W. Leishman

Brigham Young University

Abstract

Large amplitudes in a freely vibrating string cause an audible shift in normal frequencies (partials) of a piano tone due to dynamic tension changes in the string. This phenomenon is often referred to as pitch glide. This paper presents experiments that allow extraction of pitch glide by measuring both instantaneous frequencies and total energy through the use of high-speed video camera recording. Experiments were performed using a repeatable plucking mechanism on a mono-chord string apparatus. The instantaneous total energy density at a point on the string is calculated as the string's motion decays to predict the pitch glide. The prediction is verified by the instantaneous frequency measurements.

I. Introduction

When a microphone records a musical tone, a common practice to analyze the tone is to look at the frequency components present in the tone, called partials. Musical instruments produce complex waveforms, consisting of more than one partial. Figure 1 shows three examples of idealized waveforms that are similar in many respects to those a microphone would record and their corresponding spectra, from which the partials (peaks in a given spectrum) may be identified. The first row of plots corresponds to the sound radiated by an ideal tuning fork, with a sinusoidal waveform and a single partial. The second row corresponds to the idealized sound radiated by a clarinet, which has multiple partials in its spectrum. There is a fundamental partial, associated with the pitch of the tone, and higher partials, which in this case are odd multiples of the fundamental frequency. In the third row of plots, the idealized sound radiated by a trumpet is shown. The partials include both even and odd multiples of a fundamental frequency. Note here that even though the fundamental partial for the trumpet is lower in amplitude than the second partial, the pitch is nevertheless associated with the former. The relative strengths of the various partials provide important distinguishing characteristics between musical instruments when both play the same musical note (the time-amplitude envelope and the evolution of partials over time are critical distinguishing characteristics too).

The study of string vibrations is a major area of interest in the field of musical acoustics, since many musical instruments employ strings. The piano is one important example. A piano tone consists of many partials. The low-amplitude displacement solution to a freely vibrating string fixed at both ends, which yields partials that are nearly integer multiples of the fundamental partial, is well known and straightforward. These partials are stable with time. Irrespective of displacement amplitude, the partials are not exact integer multiples of the fundamental partial because of the inherent stiffness of the string (this effect is termed inharmonicity). When the string displacement amplitude becomes large enough, however, the partial frequency values change with amplitude, implying a string tension that depends on amplitude (called a dynamic tension). Thus, as the string vibration amplitude decays over time, the partial frequencies evolve accordingly. This latter effect, called pitch glide, can be audibly verified when one plucks a guitar string with significant amplitude. One will notice a decreasing pitch in the early portion of the guitar tone. As one pulls harder on a string, the tension increases, which in turn increases the wave speed in the string and results in an increase in the initial pitch of the tone. Even

moderate initial amplitudes in a vibrating string, commonly seen in musical performance, lead to an audible deviation from the low-amplitude partial frequencies.

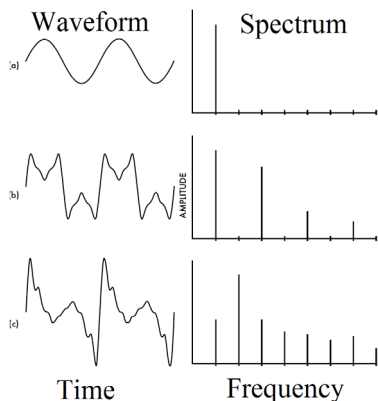


Figure 1. Idealized illustrations of the sound radiated by a tuning fork (top), a clarinet (middle), and a trumpet (bottom).

Much research has been published on improving the vibrating string model. With regard to modeling the inharmonicity of the strings, Fletcher developed a theoretical expression to predict the inharmonicity of piano string partial frequencies by including a fourth-order stiffness term in the differential equations of motion [1]. Anderson and Strong determined the effect of Fletcher's inharmonicity on the pitch of piano tones [2].

With regard specifically to the study of pitch glide, several research advancements have been made. Morse and Ingard derived coupled differential equations that govern the motion of a string in three dimensions at large amplitudes [3]. Anand then solved these equations (with damping) for the specific case of sinusoidal motion and demonstrated that any point on the string traces a slowly decaying elliptical path at a frequency proportional to a given nonlinearity parameter [4]. Bilbao, on the other hand, developed a numerical method to solve the equations that uses conservation of energy to achieve global stability for the algorithm [5]. He also mentioned that the most noticeable non-linear effect was an increase in gross propagation speed (which translates to pitch glide); however, he did not quantify this in terms of known values, such as energy. Lee et al. developed a model with re-

spect to time that predicted the frequency shift as an exponential decay, along with the decay of the sound [6]. This model was then used for piano sound synthesis. Tolonen et al. developed a nonlinear model for tension modulation in terms of string length, which relates directly to pitch glide [7].

The model of pitch glide used in this paper is not vastly different from previous models. The goal of this paper is to show how pitch glide can be related to a simple quantity that can be measured experimentally, making study of this phenomenon more accessible. A theoretical model is presented here describing the partial frequency correction due to large amplitudes, or pitch glide, with respect to the total vibrating energy in the string. The model is derived (using approximations) from the dynamic tension given in Morse and Ingard [3]. The experimental advantage of this dependence on the total energy (the total energy equals the kinetic energy of the string motion plus the potential energy due to the position of the string and the tension restoring force) is that the total energy density at a point on the string may be measured by extracting both the position and slope of the vibrating string, which yield the potential and kinetic energies, respectively [3]. Experiments were performed to extract the partial frequencies and the total energy density of a plucked string vibrating with large initial displacements as a function of time. We employed a high-speed video camera to monitor the displacement of a small portion of the string as it decays. The localized slope of the string, in the video frame window, was also extracted at the same location. Using the displacement and the slope at a point on the string, we then calculated the total energy density at that point and therefore the total energy in the string. We then compared the partial frequency values to the modified theory that we present, which depends on this total energy.

Another purpose of this paper is to introduce the instantaneous frequency method to researchers who may also be seeking a better method to track changes in frequencies over time. The instantaneous frequency (IF) method was used here to track transient partial frequencies of the plucked string, since standard fast Fourier transform (FFT) techniques yield insufficient frequency resolution over one period of the fundamental frequency. One of the first publications that introduced the concept of IF was authored by Carson and Fry in 1937 [8]. Several subsequent modifications have led to the various definitions of IF today [9-13]. The prediction and high-resolution measurement of frequency drift is extremely useful to better understand the musical aspect of stringed instruments when plucked or struck hard.

II. Theoretical model

Consider a freely vibrating string of length l , fixed at both ends (one end is located at $x = 0$) with no internal stiffness (see Fig. 2 for an illustration of the geometry). The tension in the string, T , for relatively low amplitudes, may be approximated as a constant value. It is well known that these assumptions lead to harmonic partials frequencies, v_n ,

$$v_n = \frac{nc}{2l} = \frac{n}{2l} \sqrt{\frac{T}{\mu}}, \quad (1)$$

where n is the mode number (integer values), $c = \sqrt{T/\mu}$ is the transverse wave speed, and μ is the linear mass density of the string. These partials are amplitude independent.

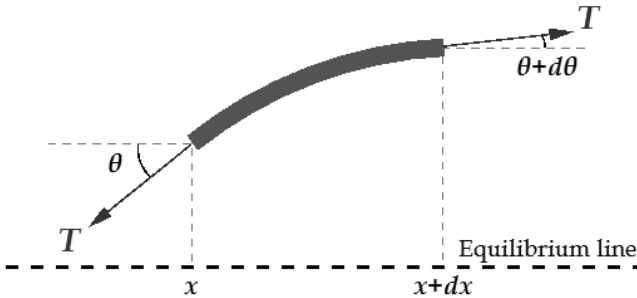


Figure 2. Illustration of a small portion of the string under tension. The y and z dimensions are perpendicular to the equilibrium line shown here.

When a string is plucked or struck with large amplitude, the tension depends on the amplitude and is therefore dynamic. Morse and Ingard [3] claim that the dynamic tension is given by

$$T(x, t) = T_0 + QA \left(\left| \frac{\partial \mathbf{R}}{\partial x} \right| - 1 \right), \quad (2)$$

where T_0 is the equilibrium tension when the string is not in motion, Q is the Young's modulus for the string material, A is the cross-sectional area of the string, and \mathbf{R} is the position vector from the origin to the point on the string that was at $(x, 0, 0)$ in equilibrium. This formula is generalized to three dimensions. The dynamic tension consists of small oscillations with a minimum value of T_0 . We approximate $|\partial\mathbf{R}/\partial x|$ to first order, assuming that longitudinal displacement is much less than transverse displacement and can be neglected. Let $\mathbf{R} = x\mathbf{e}_x + \mathbf{r}(x, t)$, where \mathbf{r} is the displacement vector from equilibrium, denoted as $\mathbf{r}(x, t) = \xi(x, t)\mathbf{e}_x + \eta(x, t)\mathbf{e}_y + \zeta(x, t)\mathbf{e}_z$. Then

$$\begin{aligned} \left| \frac{\partial\mathbf{R}}{\partial x} \right| &= \sqrt{\left(1 + \frac{\partial\xi}{\partial x}\right)^2 + \left(\frac{\partial\eta}{\partial x}\right)^2 + \left(\frac{\partial\zeta}{\partial x}\right)^2} \\ &\approx 1 + \frac{1}{2} \left(\frac{\partial\eta}{\partial x}\right)^2 + \frac{1}{2} \left(\frac{\partial\zeta}{\partial x}\right)^2, \end{aligned} \quad (3)$$

so that

$$T(x, t) = T_0 + \frac{1}{2}QA \left(\left(\frac{\partial\eta}{\partial x}\right)^2 + \left(\frac{\partial\zeta}{\partial x}\right)^2 \right), \quad (4)$$

and the averages are

$$\langle\langle T \rangle_t\rangle_x = T_0 + \frac{1}{2}QA \left(\left\langle\left\langle\left(\frac{\partial\eta}{\partial x}\right)^2\right\rangle_t\right\rangle_x + \left\langle\left\langle\left(\frac{\partial\zeta}{\partial x}\right)^2\right\rangle_t\right\rangle_x \right), \quad (5)$$

where the double angular brackets indicate both time and spatial averages. The time average is assumed to locally be valid for a few cycles, but obviously not over the entire tone. The terms $\partial\eta/\partial x$ and $\partial\zeta/\partial x$ correspond to the slopes in the y and z directions, respectively, and when squared, they are proportional to the potential energies of string deflections in those respective dimensions [14], and with the first-order

assumption that the tension (required to scale these energy terms appropriately) is the initial tension, it turns out that

$$\begin{aligned} \left\langle \left\langle \left(\frac{\partial \eta}{\partial x} \right)^2 \right\rangle \right\rangle_{t,x} + \left\langle \left\langle \left(\frac{\partial \zeta}{\partial x} \right)^2 \right\rangle \right\rangle_{t,x} &= \frac{E_y + E_z}{Tl} \\ &= \frac{E}{Tl} \approx \frac{E}{T_0 l}, \end{aligned} \quad (6)$$

where E_y and E_z are the energies for the respective dimensions of motion and E is the total energy in the string, so that

$$\langle \langle T \rangle \rangle = T_0 \left[1 + \frac{1}{2} \left(\frac{QA}{T_0} \right) \left(\frac{E}{T_0 l} \right) \right]. \quad (7)$$

Using Eq. (7) in Eq. (1), we find that large amplitudes cause the partial frequencies ν_n to deviate from their stable values $\nu_n^{(0)}$ by the ratio

$$\frac{\nu_n}{\nu_n^{(0)}} = \sqrt{1 + \varepsilon}, \quad (8)$$

where the correction parameter $\varepsilon = \frac{1}{2} (QA/T_0) (E/T_0 l)$ is proportional to the total energy in the string.

In musical acoustics, frequency deviations are often expressed in cents C , where 100 cents is equal to a semitone interval in an equally tempered 12-tone scale. The shifted frequencies then differ from the stable values, in cents, according to the relationship

$$\begin{aligned} C &= 1200 \log_2 \left(\frac{\nu_n}{\nu_n^{(0)}} \right), \\ &= 1200 \log_2 \sqrt{1 + \varepsilon}, \\ &= 600 \log_2 (1 + \varepsilon). \end{aligned} \quad (9)$$

To get an idea of how significant a change can happen in practice because of ε , let us take a look at what it would take to produce a deviation of an entire semitone from the stable partials:

$$C = 100 = 600 \log_2(1 + \varepsilon) ,$$

$$\varepsilon = 2^{1/6} - 1 \approx 0.12246 . \quad (10)$$

For a typical piano string, assuming only the first partial is present, this would imply a transverse displacement only 1% the length of the string. Thus, a significant, sometimes dramatic, effect in many applied cases can be expected.

III. Instantaneous frequency method

Conventionally, if one needed to measure the frequencies incorporated in a given signal, one would utilize the standard Fourier transform to convert the time data into the frequency domain. However, for finite digital signals, the Fourier transform has limited frequency resolution, which makes it difficult to see how a transient signal shifts in frequency over a short period of time. Other transformations or signal analyses can obtain higher resolution. One of these, the instantaneous frequency (IF) method, inherently yields near-perfect resolution in time and frequency [9], yet may be highly susceptible to bias errors from several sources, especially noise. We found that for our experiment these biases could be controlled enough to achieve good reliability along with the high resolution.

Any signal $f(t)$ may be expressed in analytic form as

$$f(t) = A(t)e^{i\phi(t)} , \quad (11)$$

where A and ϕ are both functions of the independent variable (time in this case). The instantaneous frequency of $f(t)$ is defined to be the time derivative of the phase ϕ (divided by 2π , if expressed in Hz instead of radians per second). A standard FFT of a digital signal produces an average spectrum corresponding to the information contained throughout the time length of the signal, whereas the IF of a digital signal produces a single frequency value output for each time sample in the signal. This means that to successfully implement IF for a piano tone, we must carefully employ an appropriate band-pass filter over the partial of interest to obtain a nearly sinusoidal signal.

The basic steps of obtaining the IF of a real digital signal, as given by Boashash [10], are:

1. Compute the FFT of the signal.

2. Set the negative and zero frequency (DC) components equal to zero to make the signal analytic.
3. Apply an appropriate band-pass filter around the shifting partial of interest.
4. Apply the inverse FFT to the analytic frequency-domain signal.
5. Obtain the instantaneous phase from its real and imaginary components.
6. Differentiate the phase with respect to time and divide by 2π .

Because we are especially interested in the attack portion of the recorded signal, we use a Bessel band-pass filter with the IF method to preserve the waveform of the band-limited signal and increase reliability over the attack phase at the beginning of the signal, since this portion contains the greatest correction to the frequency. Filters other than a Bessel filter did not produce physical results for the instantaneous frequency at the beginning and/or end of the time signal.

IV. Experiment

It is not necessary to measure the displacement of a freely vibrating string across its entire length to estimate its total energy. The following mathematical argument shows that having only the velocity and slope of a single point on the string is sufficient (although to obtain the slope, one must observe the neighboring spatial points), as long as the decay is not extremely rapid. Morse and Ingard [3] showed that the total energy in the string E , due to its vertical motion, is given by

$$E = \int_0^l \left(\frac{1}{2} \mu \left(\frac{\partial y}{\partial t} \right)^2 + \frac{1}{2} T \left(\frac{\partial y}{\partial x} \right)^2 \right) dx. \quad (12)$$

Before evaluating the integral in Eq. (12), we assume that the total energy density (given by the integrand) is the same over one period of oscillation as neighboring periods of oscillation. This assumption is also appropriate for the case of a slowly decaying signal, such as those measured in the experiment presented in this section, since the energy is approximately constant over the range of a few oscillations. This permits us to take a time average over the integrand. Evaluation of the integral simply produces an l (because the time averages over squared sinusoidal functions of x results in constants that are independent of x), so the total energy may be inferred from the total energy density at

any value of x , as long as it is time averaged over a period. Therefore, we are able to calculate the energy (remembering that $c^2 = T/\mu$) after measuring the slope and displacement of a very small section of the string, according to the formula

$$\frac{E}{T_0 l} \approx \frac{E}{Tl} = \frac{1}{2} \left(\frac{1}{c^2} \left\langle \left(\frac{\partial y}{\partial t} \right)^2 \right\rangle + \left\langle \left(\frac{\partial y}{\partial x} \right)^2 \right\rangle \right). \quad (13)$$

To measure both the IF and the total energy in a freely vibrating string, Rösclau piano wire was strung across a monochord apparatus and high-speed video was taken of a small portion of the string during and after the pluck. The video recording allows extraction of the position and the slope of the string, from which we can predict the total energy. Figure 3 shows a photograph of the experimental setup. The speaking length of the string (between the fixed ends) was 77.5 cm, the camera was located 7.3 cm from the left end, while the plucking mechanism was located 13.7 cm from the right end. We note here that we assumed that the exact position along the string at which these video recordings were made should not change the total energy in the string. The camera records 7.2 seconds of data at 9000 frames per second (fps) for each trial. It has a visual window of height 12 mm and width 1.5 mm (512×64 pixels, grayscale). The position scaling of the video frames was determined from the known diameter of the string (0.94 mm) in the picture. The plucking mechanism consists of a rod that slides through a tight aperture fastened to the base of the monochord apparatus. The rod (1-cm diameter) is positioned so that the string is initially stretched underneath the edge of the rod, and then it is pulled out quickly to pluck the string. The high-speed camera recording was started shortly before the pluck. The string is initially displaced 1 cm (the diameter of the rod) from equilibrium.

The Rösclau piano wire is made of steel with a density of 8 g/cm³ and a Young's modulus of 206 GPa (as reported by the manufacturer). The diameter is measured to be 0.94 mm. The string was stretched to an equilibrium tension of approximately 340 N. This last value was calculated by using Eq. (1) and measuring the low-amplitude fundamental frequency at 159.6 Hz. The string was plucked hard several times and allowed to settle over a few days to stabilize the tension before making any recordings or determining the above-mentioned physical quantities.

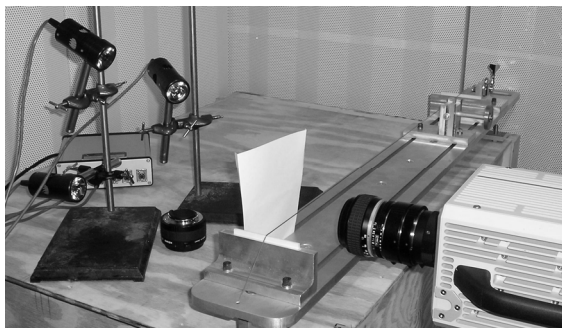


Figure 3. Experimental setup for a high-speed video recording of a plucked string on a monochord.

Six video recordings were made in total, the first four to capture vertical motion and the last two to capture horizontal motion (by laying the apparatus on its side). Both axes of motion are necessary in calculating the total energy, since some horizontal motion is inevitable. Because we cannot simultaneously record both axes of motion, we instead determined the repeatability of the pluck, so that a successive pluck could be treated as if it was the same, once the apparatus is tilted to get horizontal motion. It will be shown later that the repeatability of the pluck is very good.

The video was exported as a sequence of JPEG images and post-processed in MATLAB. The slope and position of the string in a given frame was calculated by scanning each column for a cutoff color intensity that best described the top and bottom edges of the string. The positions of these pixels were then fitted to a line, in the least-squares sense, for both the top and bottom edges separately. The displacement of the string was then chosen as the average displacement between the midpoints of the two edges of the string, and the slope was determined as the average of the edges' slopes. The displacement velocity was then calculated by using the center difference formula for the derivative of a digitized signal y ,

$$\dot{y}_n = \frac{y_{n+1} - y_{n-1}}{2\Delta t}, \quad (14)$$

where n denotes the sample number and Δt is the sampling period.

Figure 4 shows the first several frames after excitation for Test 4. As is evident from the figure, the faster string movement yields blurrier

images; however, through analysis not shown here, the edge extraction was determined to be adequately reliable for calculating the velocity and slope of the string from video.

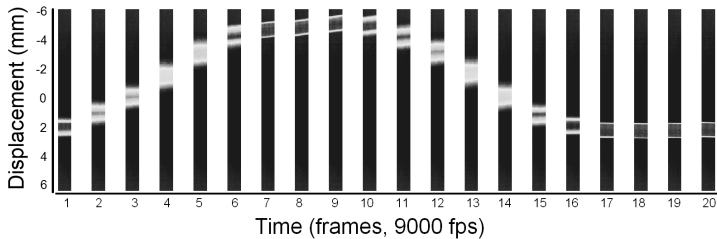


Figure 4. Frames 6511 to 6530 (listed as frames 1–20) of Test 4 video at 9000 fps, immediately after excitation of the string. Each frame measures 1.5 mm by 12 mm (although there may be some stretching distortion of the image to fit these 20 frames in this figure).

V. Results

Before comparing the experimental measurements of the frequencies to the theoretical predictions, it was first necessary to verify the repeatability of the pluck so we could add horizontal and vertical energies together, even though they were measured in different recordings. Figure 5 shows the first few oscillations of four different vertical tests (Tests 1–4) and two different horizontal tests (Tests 5–6), displaying displacement versus time for each test, in (a) and (b) respectively. The plots reveal that the different tests started out almost exactly the same, demonstrating the high degree of repeatability. A second later, the different vertical and horizontal tests were similar in shape although they were out of phase. Since we are only concerned with the energy (averaged over each period) being the same for each test for a given direction of motion, phase mismatch should not result in any significant disagreement.

We next compared the measured IF with the predicted values dependent on the total energy in the string. To calculate the total energy, we added the measured energies of Test 4 (vertical motion) and Test 5 (horizontal motion) and averaged over each 6.26566 ms (approximately one period). To obtain the measured frequency drift, we converted the absolute IFs to cents. This required an estimate of the stable frequencies. However, there is always some amount of energy in the string throughout the test, as it was not allowed to vibrate indefinitely until it

came to rest during the recording. This yields some deviation from the equilibrium frequencies in the theoretical model at low amplitudes. Thus, we chose the stable frequencies corresponding to the IFs of the partials for optimal agreement with the model from the tail end of the recorded signal.

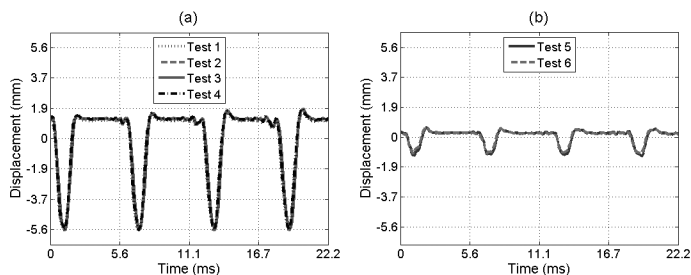


Figure 5. (a) The first few oscillations of vertical motion of the string for Tests 1–4. (b) The first few oscillations of horizontal motion of the string for Tests 5 and 6.

Figure 6(a) plots partials 1 through 3 in cents next to the theoretical model, using input from the energy measurements. The partials deviate from their stable values very similarly to each other. This is expected since, according to the theory, the deviation should not have any dependence on the partial number. At later times in the recording, below about 20 cents, the model agrees very well with the measured partial frequencies, but at higher amplitudes in the attack phase of the signal, the model overshoots by as much as 29% of the measured partial frequencies. To see directly how much the model deviates, Fig. 6(b) shows the difference in cents between the first partial and the model's prediction. Here, we can see that even in the range between about 20 and 40 cents (from 0.5 to 1 s on the time axis), the model is still only about 5 cents sharp relative to the measured values. Thus, the model predicts the change in frequency fairly accurately, except for very large vibration amplitudes.

VI. Conclusions

The theoretical model proposed in this work introduced a correction to the normal frequencies (partials) of a freely vibrating string, derived from the second-order wave equation, that is proportional to the total energy in the string. The model employed time and spatial averages of the dynamic tension, instead of the commonly used equilibrium

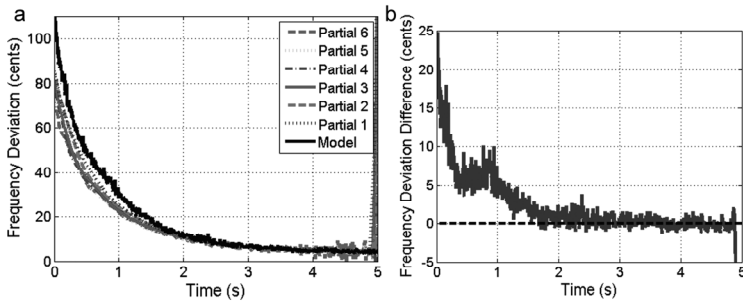


Figure 6. (a) The instantaneous frequencies (in cents) of partials 1 through 3 of Test 4 plotted against the model based on energies measured from Test 4 (vertical motion) and Test 5 (horizontal motion). (b) Difference (in cents) between the instantaneous frequency of the fundamental (first partial) and the model.

tension, and showed that these averages produce a correction term that is proportional to the total energy in the string. To test the model, a high-speed video recording was made of a vibrating string, from which the IFs of the partials were extracted (using a band-pass Bessel filter for each partial) and also the total energy of the string, both as functions of time. As long as the noise level was low (as was the case for the first several partials extracted from the video), the IF method with the Bessel filter proved to be very reliable. The total energy was calculated by extracting the slope and velocity of a small portion of the string, giving the total energy density at that point, which was then averaged over each period of oscillation. This time-averaged total energy density produced a uniform value over the entire length of the string. This method of measuring the energy thus made it possible to obtain high resolution on a very small portion of the string, rather than capturing video of the entire length of the string.

The comparison of results showed that the partial frequencies indeed are much sharper at higher displacement amplitudes, and the theoretical model accurately predicts this correction up to about 40 cents [cents are defined in Eq. (9)]. Beyond this, the model overshoots somewhat significantly. The measured IFs of the partials deviated as much as 83 cents (from their steady state values) at the beginning of the pluck, almost an entire semitone, when the string was initially displaced by only about 1 cm at the position of the plucking mechanism. Smaller initial displacements can cause a string's partial frequencies to deviate by a few cents, which is detectable by the human ear and can significantly affect the tuning of musical instruments.

More rigorous methods for extracting the energy in the string may also be necessary to more accurately form the theoretical model. Further research into the problem of a string vibrating at large amplitudes may consider the effect the dynamic tension has on the differential equation when it is not approximated as a constant. Furthermore, including Fletcher's fourth-order term in the differential equation may produce a more accurate model. These investigations could lead to higher-order corrections to the frequencies that are valid at very large amplitudes. Finally, further research into the effect of mode coupling and multiple-string coupling on normal frequencies may prove to be useful in application to real musical instruments.

Acknowledgments

This research was made possible by funding from the BYU Department of Physics and Astronomy. We appreciate Scott Thomson allowing us to use his high-speed camera. Some advice and technical assistance was provided by William Strong and Brian Thornock.

References

- [1] H. Fletcher, "Normal vibration frequencies of a stiff piano string," *J. Acoust. Soc. Am.* **36(1)**, 203-209 (1964).
- [2] B.E. Anderson and W.J. Strong, "The effect on pitch due to the inharmonicity of piano tones," *J. Acoust. Soc. Am.* **117(5)**, 3268-3272 (2005).
- [3] P.M. Morse and K.U. Ingard, *Theoretical Acoustics*, (New York, McGraw Hill, 1968), p. 856-859.
- [4] G.V. Anand, "Large-amplitude damped free vibration of a stretched string," *J. Acoust. Soc. Am.* **45(5)**, 1089-1096 (1969).
- [5] S. Bilbao, "Conservative numerical methods for nonlinear strings," *J. Acoust. Soc. Am.* **118(5)**, 3316-3327 (2005).
- [6] N. Lee, J. O. Smith III, J. Abel and D. Berners, "Pitch glide analysis and synthesis from recorded tones." *Proc. of the 12th Int. Conference on Digital Audio Effects (DAFx-09)*, Como, Italy, September 1-4, 2009.

- [7] T. Tolonen, V. Välimäki and M. Karjalainen, "Modeling of tension modulation nonlinearity in plucked strings." IEEE Transactions on Speech and Audio Processing, **8(3)**, May 2000.
- [8] J. Carson and T. Fry, "Variable frequency electric circuit theory with application to the theory of frequency modulation," *Bell System Tech. J.*, vol. 16, pp. 513-540, 1937.
- [9] J. Ville, "Theorie et application de la notion de signal analytic," *Cables et Transmissions*, vol. 2A(1), pp. 61-74, Paris, France, 1948. Translation by I. Selin, "Theory and applications of the notion of complex signal," Report T-92, RAND Corporation, Santa Monica, CA.
- [10] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal—Part 1: Fundamentals," Proc. IEEE **80(4)**, 520-538 (1992).
- [11] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal—Part 2: Algorithms and Applications," Proc. IEEE **80(4)**, 540-568 (1992).
- [12] H. Suzuki, F. Ma, H. Izumi, O. Yamazaki, S. Okawa and K. Kido, "Instantaneous frequencies of signals obtained by the analytic signal method," *Acoust. Sci. & Tech.* **27(3)**, 163-170 (2006).
- [13] L. Rossi and G. Girolami, "Instantaneous frequency and short term Fourier transforms: Application to piano sounds," *J. Acoust. Soc. Am.* **110(5)**, 2412-2420 (2001).
- [14] L.E. Kinsler, A.R. Frey, A.B. Cripps, J.V. Sanders, *Fundamentals of Acoustics 4th Ed.*, (New York, John Wiley & Sons, Inc, 2000), p. 58-59.

Fabrication and Testing of a Strain-Based Carbon Nanotube Magnetometer Structure*

Jonathon A. Brame

Brigham Young University

Johnathan E. Goodsell

Brigham Young University

Stephanie. A. Getty

NASA

Y. Zheng

NASA

David D. Allred

Brigham Young University

* This paper was presented at the 2010 conference of the Utah Academy of Sciences, Arts, and Letters.

ABSTRACT

A test structure for a prototype nanoscale magnetometer exploiting the strain sensitivity of single-walled carbon nanotubes (SWCNTs) has been fabricated. The nanotube magnetometer would boast reduced dimensions, mass, and power requirements compared with a Fluxgate magnetometer. Dramatic resistance increase with strain has been previously reported for individual nanotubes, and this magnetometer design concept seeks to extend this strain-resistance property to an "as-grown" ensemble of SWCNTs. Measurements of a test structure show a correlation between applied magnetic field and device conductivity. This correlation indicates an increase in conductivity with strain to the network of nanotubes; candidate mechanisms for this behavior are discussed.

INTRODUCTION

A magnetometer is an instrument for measuring magnetic fields. There are two basic categories of magnetometers: scalar magnetometers, which measure the magnitude of a magnetic field, and vector magnetometers, which measure the strength along a known direction of a magnetic field. Some magnetometers can measure magnetic fields below one nanotesla. The objective of this work is to develop a new vector magnetometer technology that operates in a wide range of temperatures with lower power consumption, within a smaller footprint, and with higher spatial resolution than conventional magnetometer technologies, such as Fluxgate or SQUID magnetometers.

Space exploration provides a wide range of uses for magnetometer devices. Magnetometers are used for scientific measurement, such as mapping the interaction of solar wind with planetary magnetospheres and probing the interior of planets. In planetary exploration, magnetometers can provide large- and small-scale geologic mapping and are critical for orienting spacecraft and instruments. In the future, nanoscale magnetometers may prove useful for astronaut and rover orientation.

Because of their remarkable electronic and mechanical properties, carbon nanotubes, particularly single-walled carbon nanotubes (SWCNTs), have been the subject of intensive research efforts over the past decade. The bonding structure of carbon nanotubes is predominantly sp^2 , as in graphite. SWCNTs can be thought of as a thin strip of graphene (single-layer graphite) rolled up to form a seamless tube.¹ The aspect ratio (ratio of length-to-width) of these tubes can be $>10^6$, and SWCNTs are stronger per weight than steel.

An additional property, the focus of this investigation, is that as the tubes are stretched or bent their electrical resistance increases by orders of magnitude for very small strains.¹⁻⁵ Many researchers have already suggested that this large strain-induced response in electronic resistance, or piezoresistance, can be exploited in strain-sensing devices.⁶⁻⁸

An as-grown ensemble of SWCNTs was selected to develop one such strain-sensing device for magnetic field detection. Although a SWCNT ensemble will be less sensitive to strain than an isolated SWCNT, a network of SWCNTs is used to promote device uniformity and ease of device integration.

The proposed magnetometer design consists of a microscale iron needle suspended over a trench and mechanically coupled to a mat of electrically contacted nanotubes¹⁰ (see Figure 1). In the presence of a magnetic field, the iron needle will torque to align with the field similar to a traditional compass. This torque will strain the nanotubes on which the needle is suspended, thereby changing their resistance. Upon calibration, this change in resistance will correspond to a magnetic field strength. Device dimensions would allow for 10,000 devices to fit onto a 1×1-cm instrument, enabling high spatial resolution of both magnetic field strength and direction.

The theoretical capability of this SWCNT-based magnetometer compares favorably with currently available devices for space applications, such as the Fluxgate magnetometer, magnetoresistive devices (MR and anisotropic magnetoresistor, AMR), and an optically pumped scalar instrument. Compared with these instruments, the proposed SWCNT magnetometer would operate at lower power, occupy a smaller footprint, and permit orders-of-magnitude higher spatial resolution, which may prove useful in studies of the magnetic microstructure on other planets.

A 2×2 array of proof-of-concept magnetometer test structures has been successfully fabricated. This paper details the fabrication process for the array of magnetometer test structures. Magnetic field testing was performed on one successful test structure to determine the preliminary strain-resistance response of the prototypes. Initial test results are presented and discussed. Finally, future refinements to the next iteration of magnetometer test structures are outlined.

METHODS

A schematic of the test structure processing is shown in Figure 1. An ensemble of SWCNTs is grown on a Si/SiO₂ wafer. Gold contact pads are deposited on the wafer surface, followed by an iron needle

between the gold pads. Finally, a trench is etched under the needle, which is supported by the SWCNTs, allowing the needle to rotate and align itself with a magnetic field. Details for each processing step follow.

The substrate was a 1×1 -cm (100) silicon chip cut from a 500- μm -thick wafer with a 500-nm oxide capping layer. A dense ensemble growth of SWCNTs was achieved using an indirect catalyst evaporation⁸⁻¹³ and chemical vapor deposition (CVD).

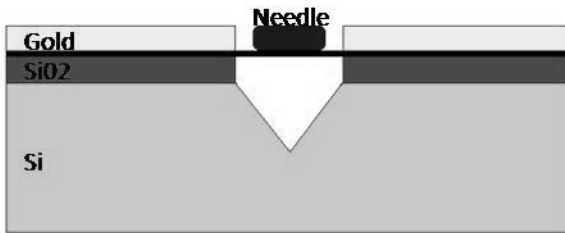


Figure 1. Cross-section view of device schematic showing the edge of the gold needle (pointing perpendicular to page) suspended on the nanotubes, which are contacted by the gold pads on either side of the V-shaped trench. The dark line beneath the gold layer and holding the needle represents the nanotube mat.

Patterns for the contact pads were generated using electron-beam lithography (EBL). The sample was coated with a bilayer electron-beam resist (EBR) of methyl methacrylate/methacrylic acid (MMA/MAA) followed by polymethyl methacrylate (PMMA). The bilayer resist is commonly used in semiconductor processing to provide an undercut to facilitate removal of metal by liftoff outside the desired pattern area. The resist-coated sample was placed in a scanning electron microscope (SEM) fitted with a Nanometer Pattern Generation System. Four identical sets of gold contact patterns were written on the sample. The exposed sample was developed by sequential dipping in a solution of toluene:isopropyl alcohol (IPA) (1:3) for 5 seconds, IPA rinse for 10 seconds, 2-ethoxyethyl acetate:ethanol (2:5) for 20 seconds, and IPA rinse for 10 seconds.

Gold at a thickness of 100 nm was evaporated on the sample to form the contact pads and was preceded by a 10-nm-thick chromium underlayer to promote surface adhesion. The evaporator base pressure was 1×10^{-6} torr, and two resistive evaporation sources were used to

avoid venting the chamber in between evaporations. Liftoff of the excess metal was performed by placing the gold-coated sample in acetone for approximately 2 hours with periodic agitation to remove the EBR and unwanted metal. Liftoff yielded four sets of patterned gold electrical contact pads with alignment marks atop a SWCNT mat.

The needles were patterned by spin casting another layer of bi-layer EBR and exposing four $80 \times 3\text{-}\mu\text{m}$ needle patterns using the EBL process as above with an alignment step, such that each needle was centered between the gold electrodes. The sample was developed following the process described above.

The needle evaporation consisted of a 30-nm layer of chromium, followed by a 300-nm layer of iron, capped by another 30-nm layer of chromium. The chromium layers are used to passivate the iron against oxidation. Liftoff was again performed as above and yielded chromium/iron/chromium needles centered between gold contact pads atop the mat of SWCNTs (see Figure 2).

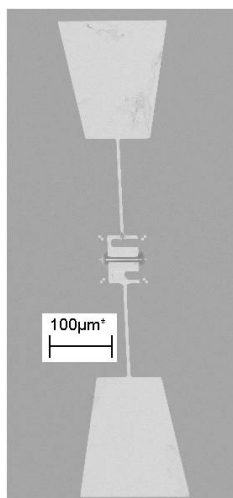


Figure 2. Scanning electron microscope image of entire device. The large trapezoidal structures at the top and bottom are the gold contact pads, with the trench and needle in the middle. The trench is the dark region surrounding the needle. The nanotubes are not visible but are contacted by the C-shaped pad above and below the needle. The four small patterns near the outside of the C-shapes are alignment marks.

The trenches beneath the needles were patterned using photolithography. 1812 photoresist (PR) was applied to the sample and soft

baked at 110°C for 10 minutes. An optical microscope was then used to expose oval-shaped patterns in the PR over each needle. To form each oval-shaped exposure, four overlapping circles were exposed for 45 seconds at 95% illumination intensity using the microscope's aperture at 1000× magnification. These exposures yielded an oval-shaped trench pattern of approximate dimensions 35×90 μm . The sample was then developed for 1 minute and then was hard baked at 110°C for 90 minutes.

The trench was formed by etching the oxide layer in buffered oxide etch (BOE) for 20 minutes with slight agitation, where the patterned photoresist acts as an etch stop, followed by a 20-minute rinse in deionized water. BOE is known to isotropically etch through the silicon dioxide surface layer. With the patterned oxide as an etch stop, the sample was then etched for 2 hours in agitated KOH at 65°C. KOH anisotropically etches through silicon and forms a pyramidal trench pattern in (100)-oriented wafers. Notably, the chromium/iron/chromium metal was found to be chemically robust to this wet etch.

The etched sample was finally submerged in IPA in a critical point dryer (CPD). The CPD mitigated the potentially damaging surface tension effects that could cause the needle to adhere to the trench bottom during normal evaporative drying. Scanning electron microscopy was used to examine the devices after processing (see Figure 2-3).

Adhesion between the needle and SWCNT network was found to be a point of failure during agitated etching of the trench; as a result, the needles were often found to be partially removed after trench formation. For magnetic testing, the most complete structure was selected to measure the electric current across the SWCNTs as a magnetic field was applied perpendicular to the SWCNT mat. The two gold contact pads were wire bonded to a chip carrier and connected to a probe station. Current across the contacts (device current) was measured using a Stanford Research Systems SR570 low-noise current pre-amplifier as a voltage of 30 mV was applied across the contacts using a Sorensen XT15-4 power supply, resistively divided using an electrostatic discharge-mitigating electronic circuit.

The test magnetic field was generated by a GMW 5201 Projected Field Electromagnet, and the sample was oriented such that the magnetic field lines passed perpendicularly through the SWCNT mat, which would tend to rotate the iron needle out of the SWCNT plane. The magnetic field was incrementally applied by increasing the electromagnet current in 1 A steps up to a maximum of 15 A. This corresponds to a magnetic-field step of 0.0267 T, yielding a 0.4-T maximum

field, after which the field was stepwise decreased. Each step lasted 1 minute to eliminate transient effects.

RESULTS

An array of four magnetometer test structures were successfully fabricated. SEM investigation revealed that the SWCNT ensembles, the gold contact pads, and the iron needles were chemically robust to the etching processes. Further, the needles were successfully suspended atop the SWCNT ensembles above trenches etched through the silicon. A successful test structure is shown in Figure 3.

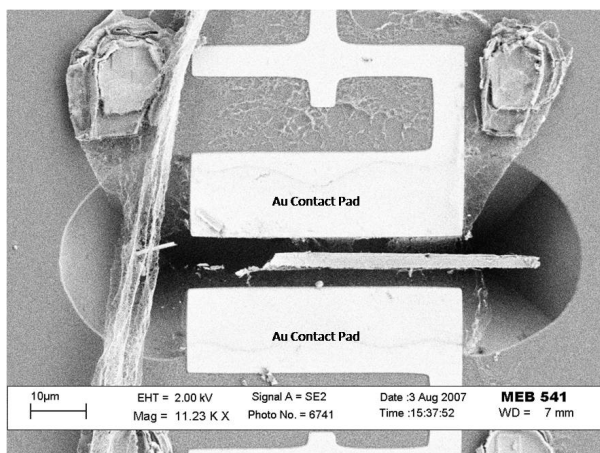


Figure 3. SEM image of the test structure, consisting of two gold pads making electrical contact with a SWCNT mat (the spider-web-type structure). An iron needle was suspended on the mat above a trench etched through the SiO_2 layer and into the silicon layer. Portions of the nanotube mat on either side of the device have delaminated from the SiO_2 layer leaving a substrate bare of nanotubes on either side of the photo. A portion on the mat of the left side has rolled up around the alignment mark and continues across the trench and is on top of the lower left alignment mark.

Representative magnetic field measurements are shown in Figure 4. Magnetic field was increased for 500 seconds and then decreased back to baseline. Initial increase in magnetic field caused an increase in device current. Device breakdown at around time 700 seconds prohibits further conclusions. An increase in device current means that the ensemble resistance of the device was decreasing with applied magnetic

field. Other tests showed greater noise with lower signal and suggest refinements to the electrode, needle geometry and to the nanotube density, discussed below.

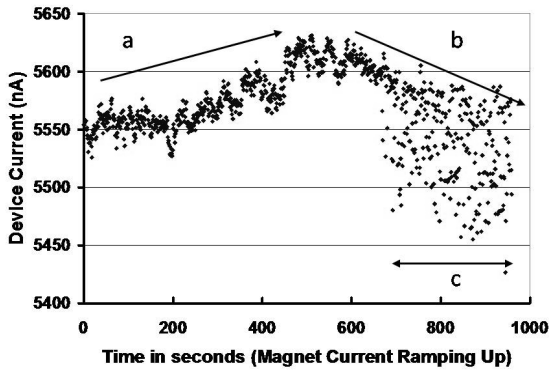


Figure 4. Device current increases as magnetic field is increased, labeled “a” on the graph. 500 seconds into the experiment the magnetic field is equal to 0.4 T, at which point the field is ramped down (“b” on graph). At some point, contact with one of the pads becomes intermittent, leading to the noisy signal in region “c” on the graph. The x-axis (time) also represents the magnetic field, which increased in one-minute intervals up to a maximum at time 500 seconds, then decreased in one-minute intervals back to zero.

According to the work of Tombler and others,¹⁻⁵ a decrease in nanotube current with an increase in magnetic field is expected. As the magnetic field torques the iron needle, the nanotubes are strained and the conductivity of the nanotube mat should decrease. In contrast to this, the current through the nanotube network increased when exposed to a magnetic field. Carbon nanotubes themselves should not display any magnetically induced conductivity changes. Further, previous work suggests that a magnetic field does not change the conductivity of a SWCNT mat grown from an iron catalyst.¹⁴ Hence, any change in the conductivity of the nanotubes is likely due to the magnetic induced torque on the iron needle.

The data obtained through magnetic testing suggest several mechanisms to explain these results and possibilities for refinements to the fabrication process. It is possible that the noisy signal is due to the fact that while the iron needle torques out-of-plane and strains the outer SWCNTs, the center SWCNTs are not significantly strained and thus provide a low-resistance current path. This could be avoided by sepa-

rately contacting the outer nanotubes using a device geometry such as suggested in Figure 5. There may also be an insufficient magnetic moment in the iron needle to torque the stiff nanotube network. By decreasing the nanotube density, the device sensitivity should increase. This will allow a more precise measurement of resistance versus magnetic field. Additionally, changing contact resistance or modulation of the number of conducting channels cannot be ruled out. Within the data available, it is difficult to refine the mechanism responsible for the unexpected increase in device conductance due to the limited data collected before device failure. Additional devices should be fabricated to confirm the reported effects. Nevertheless, the response shown in Figure 4 is suggestive of magnetic field response, and further investigations may reveal its operation.

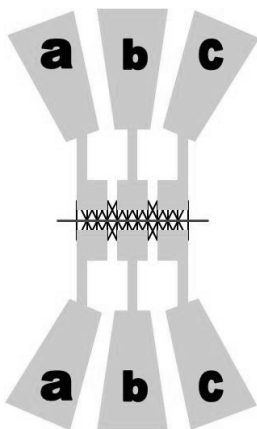


Figure 5. The design for the next iteration of gold-pad geometry features three separate sets of pads (a, b, c) to contact specific sections of nanotubes. The black Xs represents the nanotube mat, and the horizontal line is the needle. Dividing the pads into three sets makes it possible to use the outer sets of pads (a and c) to measure the strain across the outer portions of the nanotubes. These should be the most strained.

CONCLUSIONS

A fabrication process for magnetometer test structures has been demonstrated. An array of prototype magnetometer test structures has been successfully fabricated. The test structures consist of a high aspect-ratio iron needle suspended above a trench by a mat of SWCNT. Gold pads electrically contact the SWCNT mat. This fabrication proc-

ess is now available and being used in various field sensing applications.

Initial magnetic testing indicates that there is some correlation between magnetic field and measured resistance in the device. Further device fabrication and testing is necessary to establish the extent of this relationship. In addition, other methods for detecting the movement of the needle such as by changes in capacitance might be explored. Brooks et al.¹⁵ described a compact instrument for measuring the magnetization of a small sample at high field using changes in capacitance as the small sample twists in an applied magnetic field. In the place of our iron needle, they suspended their material to be characterized.

ACKNOWLEDGMENTS

We gratefully acknowledge financial support from NASA GSFC Director's Discretionary Fund, the NASA Internal Research and Development Fund, the Rocky Mountain Space Grant Consortium, and the Student/Faculty Internship Program for the summers of 2006 and 2007, which provided support for J.A. Brame, J.E. Goodsell, and D.D. Allred. The authors also acknowledge technical assistance from D. Dove, D. Stewart, and B. Munoz and numerous helpful discussions with B. Hicks, R.B. Davis, R. Vanfleet, and D. Rowland. We also acknowledge a reviewer for directing our attention to reference 15.

REFERENCES

- ¹H. Dai, "Carbon nanotubes: Synthesis, integration, and properties." *Acc. Chem. Res.* **35**, 1035-1044 (2002).
- ²T. Tombler, C. Zhou, L. Alexseyev, J. Kong, H. Dai, "Reversible electromechanical characteristics of carbon nanotubes under local-probe manipulation" *Nature* **405**, 769 (2000).
- ³A. Maiti, A. Svizhenko, and M.P. Anantram, "Electronic transport through carbon nanotubes: effects of structural deformation and tube chirality," *Phys. Rev. Lett.* **88**, 126805 (2002).
- ⁴J. Cao, Q. Wang, and H. Dai, "Electromechanical properties of metallic, quasimetallic, and semiconducting carbon nanotubes under stretching," *Phys. Rev. Lett.* **90**, 157601 (2003).
- ⁵C. Stampfer, A. Jungen, C. Hierold, "Fabrication of Single-Walled

Carbon-Nanotube-Based Pressure Sensors,” *Nano Letters* **6**, 233 (2006).

⁶Y. Su, A.G R. Evans, and A. Brunnschweiler, “Micromachined silicon cantilever paddles with piezoresistive readout for flow sensing,” *J. Micromech. Microeng.* **6**, 69 (1996).

⁷P. Dharap, Z. Li, S. Nagarajaiah, E.V. Barrera, “Nanotube film based on single-wall carbon nanotubes for strain sensing,” *Nanotechnology* **15**, 379 (2004).

⁸J.H. Hafner, M.J. Bronikowski, B.R. Azamian, P. Nikolaev, “Catalytic growth of single-wall carbon nanotubes from metal particles,” *Chem. Phys. Lett.* **296**, 195 (1998).

⁹J. Kong, H.T. Soh, A.M. Cassell, C.F. Quate, H. Dai, “Synthesis of individual single-walled carbon nanotubes on patterned silicon wafers,” *Nature* **395**, 878 (1998).

¹⁰J. Brame, SA Getty, J. Goodsell, DD Allred, “Strain-based electrical properties of systems of carbon nanotubes embedded in parylene,” *MRS Fall 2006 Conference Proceedings, Symposium Q*.

¹¹J. Goodsell, SA Getty, J. Brame, DD Allred, “Thin-film iron-catalyzed ‘beads on a string’ carbon nanotubes,” *J. Utah Acad. Sci., Arts Letts.* **84**, 130 (2008)

¹²R.G. Lacerda, K.B.K. Teo, A.S. Teh, M.H. Yang, S.H. Dalal, “Thin-film metal catalyst for the production of multi-wall and single-wall carbon nanotubes,” *J. Appl. Phys.* **96**, 4456 (2004).

¹³K. Hata, D.N. Futaba, K. Mizuno, T. Namai, M. Yumura, “Water-assisted highly efficient synthesis of impurity-free single-walled carbon nanotubes,” *Science* **306**, 1362 (2004).

¹⁴S.A. Getty, and G. Kletetschka, “Single walled carbon nanotubes for a strain-based magnetometer,” *IEEE-NANO 2006. Sixth IEEE Conference on Nanotechnology*, **2**,. 465-468, (2006).

¹⁵J.S. Brooks, M.J. Naughton, Y.P. Ma, P.M. Chaikin, and R.V. Chamberlin, “Small sample magnetometers for simultaneous magnetic and resistive measurements at low temperatures and high magnetic fields,” *Rev. Sci. Instr.* **58**, 117-121, (1987).

Parallel Computer Simulations of Dynamical Relaxation Processes and Clusters of Galaxies

Brandon K. Wiggins, Anthony R. Thomas, and Brent A. Sorensen

Southern Utah University

Abstract

Many clusters of galaxies have yet to achieve dynamical equilibrium as evidenced in the complex non-Gaussian radial velocity distributions of their cluster galaxies. Although some profiles have been identified as evidence of particular dynamical processes, many profiles remain largely unexplained. In this preliminary study, we initially follow in the footsteps of Schindler and Böhringer in simulating collisions between simple clusters of self-gravitating particles and examining the resulting radial velocity distributions in the hope of establishing dependable trends that might identify observed cluster profiles as evidence of merging between smaller systems. With greatly increased particle numbers and an improved softening algorithm, we verify the presence of skew and bimodality in radial velocity profiles in clusters merging collisionlessly along the line of sight as seen in Schindler and Böhringer. We provide results for more complex scenarios involving three initially equilibrated clusters. Finally, we discuss these results in the context of

the profiles observed in galaxy clusters and attempt to reproduce the radial velocity profile of the entirety of the Virgo Cluster.

1. INTRODUCTION

The formation and evolution of clusters of galaxies is a topic of special interest in cosmology. Clusters of galaxies are among the largest gravitationally bound structures in the universe and contain special clues regarding its evolution. By studying certain observables of these clusters of galaxies, it is sometimes possible to identify processes or events which may be a part of a given cluster's history.

In this paper, we consider what are known as the radial velocity distributions of clusters of galaxies. The radial velocity of a galaxy is the component of a galaxy's speed along the line of sight. Observationally, each cluster member's radial velocity may be determined through the galaxy's Doppler shift. Radial velocity distributions of clusters of galaxies are constructed by collecting the radial velocities of all galaxies in a cluster and representing them in a type of histogram. In systems that have settled to a type of steady-state configuration (systems that we say have achieved dynamical equilibrium or have *relaxed*), we expect the radial velocities to be normally distributed. Clusters with radial velocities that do not appear to be normally distributed are generally interpreted to still be in the process formation (see, for example, [2] and [5]). Despite the attention many such clusters have received, the actual processes that might be producing many radial velocity distributions remain uncertain.

In this study, we seek to simulate and track the evolution of simple gravitationally bound systems as we explore the hypothesis that some observed complex profiles might be evidence of the relatively simple scenario of cluster mergers with an observer positioned with his line of sight along an axis of collision. This problem was well-treated in the literature by Schindler and Böhringer [9] though these authors performed their study with very limited computing resources. It should also be noted that the topic of collisions between clusters was not the exclusive focus of these Schindler and Böhringer's study. Advances in computers during the last 17 years has provided the opportunity to explore this question with higher levels of sophistication and detail, in addition to opening doors for further investigating possible applications for these principles in nature.

2. METHOD

Clusters of galaxies possess three major mass constituents: invisible *dark matter*, luminous matter collected in *galaxies*, and a hot diffuse gas called the *intracluster medium*. The scale and complex origin of these systems would make the task of conceiving a fully descriptive simulation of galaxy cluster dynamics and evolution very formidable. It should be noted that simulations of phenomenal sophistication exist in the literature (see, for example, [11]), which the reader is invited to review. For the scope of this study, we shall consider the results of very simple simulations and scenarios and seek to find applicable correlations between our simulation results and the radial velocity profiles observed in clusters of galaxies.

The difficulty in modeling self-gravitating systems rests in the fact that (with the exception of a few contrived examples) no closed form solutions exist for the orbits of members of a system containing more than two members. One must study these systems using computers through a numerical simulation method known as an *n-body code*. Such simulations involve advancing each particle based on the net force exerted upon it by every other particle in the system. Tracking evolution of systems containing many thousands or even millions of particles can be mammoth undertakings, requiring high levels of sophistication in computer code writing and implementation. We can simplify the problem by choosing to simulate *collisionless* gravitating systems or systems in which close encounters between particles are not considered important.

Assuming galaxy clusters behave like collisionless self-gravitating systems requires proper justification. If galaxies move about in a gravitational well largely produced by a cluster-wide dark matter halo, a collisionless system might not be too far removed from actual galaxy clusters as Cold Dark Matter, which would dominate gravitational interactions is thought to behave collisionlessly [3]. It should also be noted that collisionless n-body codes only provide good results over relatively short time scales. Fortunately, many clusters of galaxies (especially those with radial velocities which are not normally distributed) are observed today in the very early stages of their formation, stages which collisionless codes are thought to describe very well. These epochs of cluster development are dominated by a process known as *violent relaxation* (one should picture colliding, or seething, systems of particles), which is a process that such codes are best suited to study and is of principal interest in this paper.

We wrote an n-body code in the C programming language that we adapted for parallel supercomputing using NVIDIA CUDA extensions.

To compute the total force acting on each particle, we applied a simple Direct Summation (Brute Force) method with a softening kernel proposed in [3]. Force computations were executed in parallel. After the net force on each particle was determined, particles were advanced in position and velocity space using an integration scheme similar to a leapfrog algorithm. The process of computing forces and then advancing particles was repeated for each time step. The reader is invited to review the Appendix for further algorithm and implementation details.

The initial conditions for our simulations involved equilibrated clusters of equal-mass particles corresponding to the distribution function for a Plummer Sphere (see Appendix) strictly because of this potential's simplicity. Although the matter distribution in an equilibrated cluster of galaxies is thought to better follow a more complex distribution, a Plummer Sphere will provide a sufficient approximation for the scope of the current study. Algorithms for generating these initial conditions were adapted from [6]. For simplicity and generality, all simulation results were returned in n-body units (the gravitational constant and the mass of a given gravitationally bound system are equal to one unless explicitly specified). Following each simulation, particle positions and the line-of-sight or radial velocity profile along a collision axis were collected for analysis.

Schindler and Böhringer [9] have collisionlessly modeled such clusters using 500 particles with each particle representing a galaxy. Even though the galaxies provide the only observable radial velocities in a galaxy cluster, limiting the number of particles to the number of galaxies is not the most accurate means of studying the evolution of merging clusters of galaxies in the collisionless paradigm. In [9], the authors have assumed the dark matter of the cluster is largely collected in galaxies, neglecting the influence of a cluster-wide dark matter halo. This shortcoming is troubling in light of the fact that collisionless codes mimic the behavior of systems that theoretically contain an *infinite* number of particles or a continuous mass distribution, more representative of dark matter than discrete galaxies. The final realizations of collisionless n-body simulations should not be thought to represent the final positions of particular galaxies but should instead be regarded as a "sampling of the probability-density distribution of position and velocity" of the system in question [3]. An intuitive method of increasing the accuracy and descriptive power of the simulation would thus be to greatly increase the number of particles to analyze the merging process and evolution of these distributions in greater detail. Increased particle counts are further advantageous in boosting the integrity of the simulation over longer time intervals and provide results more robust to minor variations in initial conditions. Particle numbers in our simulations vary

between 7,500 and 10,000. The figures that follow include particle counts for each respective simulation.

As previously mentioned, cluster galaxies provide the only observable radial velocities for a cluster. In contrast, our simulations will be returning radial velocities for every particle in the system or the radial velocities representative of *all* mass constituents within the cluster. Although the increased particle numbers contribute greatly to the realism of the short-term behavior of matter in the merging clusters, we might properly ask ourselves how the radial velocity distributions of our simulations of thousands of particles would resemble the radial velocity profile that we would observe from only hundreds of cluster galaxies. It has been shown in the literature [4] that galaxies closely trace out background matter distributions within clusters of galaxies. For the purposes of this study, we shall assume that the positions and radial velocities of the galaxies are realized from the same positions and radial velocity distributions as the rest of the mass in a cluster of galaxies. Our simulation thus omits processes including merging between galaxies or other intergalactic interaction as galaxies are *not* directly represented by particles. The positions and velocities of particles in our simulations possess more of a statistical meaning and the distribution of radial velocities from our simulations are best thought of as probability density distributions for mass within the cluster which cluster galaxies will also be assumed to follow. It should also be noted that interactions between galaxies are not important in the global evolution of the system over the time scales we are examining. Superficial comparisons between our results and observation will be carried out by investigating similarities in the shapes of the observed and simulated distributions.

During the simulation process, we require a method to determine the physical accuracy of our simulation. The total energy of the particles in each simulation is

$$E_{\text{total}} = T + W \quad (2.1)$$

where T is the total kinetic energy of the system and W is the total potential energy. Because our system is isolated, we expect that energy is conserved. The softening kernel mentioned above was implemented to aid in energy conservation. By carefully fixing a softening length ϵ (see Appendix) and refining the step size (generally 0.005 in n-body time units), the energy error of each simulation (with the sole exception of Fig. 2) was kept within 1 percent.

3. RESULTS

We begin by considering two clusters of 5,000 particles each (see Figure 1). Each cluster has a mass of unity. One cluster is given an initial velocity of 0.5 (speeds appear in n-body units) in the direction of the other to speed the merging process. The observer is indicated by the eye graphic in Figure 1, a and b. Before the merger, the radial velocity profile appears normally distributed. At the moment of merging, the distribution spreads and gains twin peaks, one peak corresponding to each cluster as they are accelerated toward each other. We may compare and contrast this scenario to that contrived in Figure 2, where two initially equilibrated clusters are again set on a collision course, though one cluster now has a mass equal to half that of its partner. The

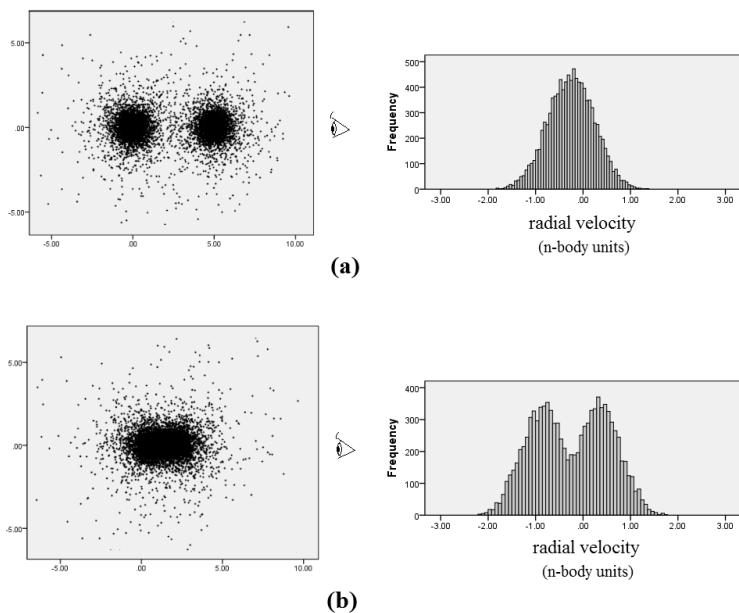


Figure 1: Spatial positions of particles appear at left and radial velocity distributions at right. In (a) we see two equilibrated Plummer Spheres. One sphere is given an initial velocity in the direction of the other. In (b), we see the two clusters right before the point of merger when a symmetric double-peaked profile appears. Simulation contained 10,000 particles and an energy error less than 1%. Eye graphics indicate the line of sight.

radial velocity profile appears slightly skewed in the initial conditions (stage 1 in Fig. 2) because of the initial velocity of 0.5 we have imparted to the smaller cluster. In this scenario, we see again that the distribution spreads in the moments directly before the collision, producing two peaks (stage 2 in Fig. 2); however the cluster with a smaller particle membership produces a peak that is smaller than its partners, effectively revealing the source of the skew in the first frame: that a smaller underlying distribution had a mean radial velocity slightly displaced from its larger partner's. As the smaller cluster is accelerated into the larger, the velocity distribution of the smaller is increasingly displaced from the larger until two distinct peaks are visible. We also note that the radial velocity profile returns quickly to what appears to be a nearly Normal distribution in stages 3 and 4 of Figure 2.

Successive simulations were carried out to discover that as greater initial velocities were imparted to the smaller cluster, the more pronounced and displaced the peaks became. Simulations with lower initial velocities yielded only skewed profiles. We conclude that the appearance of double peaks in a merging scenario as we have contrived would only be manifest if the clusters of interest were moving at sufficiently high relative speeds along the line of sight.

In Figure 3, we consider a more complicated scenario where two less massive clusters merge on perpendicular axes with a cluster with mass twice that of each the others. An observer is indicated with a graphic of an eye. As the clusters approach each other in Figure 3a, we see the characteristic skew in the radial velocity profile. Once again, however, at the point of merging in Figure 3b, the clusters are moving sufficiently fast to create distinct peaks. The peaks appear even more pronounced and separated than those in Figure 2 as the cluster moving along the line of sight is accelerated to higher speeds toward both the large cluster and the its similarly-sized sister cluster. This demonstrates (among other things) the intuitive fact that the collisions of systems with more mass would permit clusters to be moving at greater speeds. We also note, however, that although the peaks of the three clusters can be found (two peaks appear very close together in the larger spike in Figure 3b), we essentially observe two peaks, indicating that mergers with clusters perpendicular to the line of sight could not be readily discerned from a radial velocity profile.

Determining whether sufficient relative velocities between clusters could conceivably occur so that multiple peaks within the distribution are evident to an observer could be the subject of further investigation. The possibility of clusters moving at such relative speeds would have to fit within the current paradigm of cosmological large-scale structure formation.

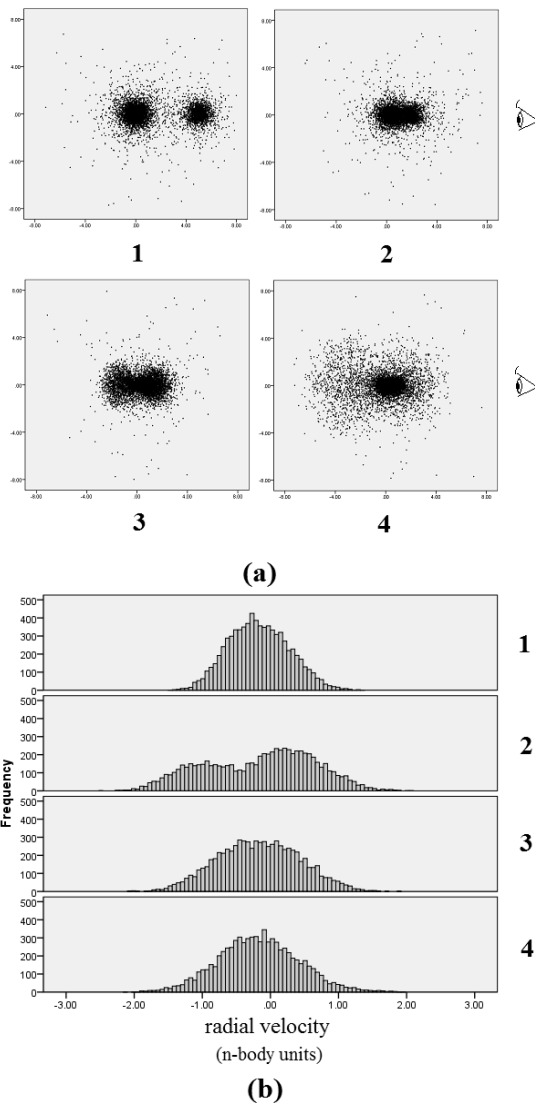


Figure 2: Simulation of merging clusters (one's mass double the others). Spatial positions of particles appear in (a) and radial velocity profiles in (b) for four times during the merger. This simulation included 7,500 particles over the course of 20,000 time steps with less than 2% energy error. Eye graphics indicate the line of sight.

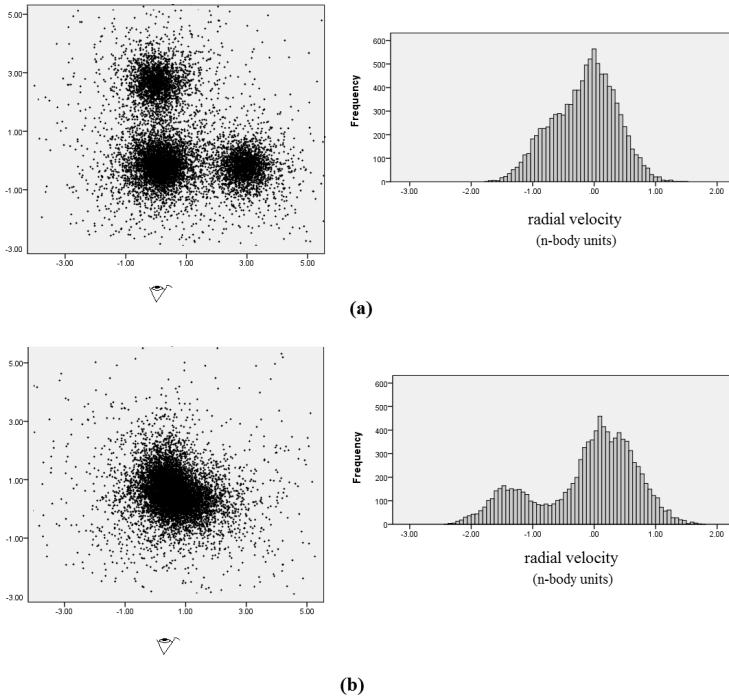


Figure 3. Simulation of a large cluster merging with two smaller clusters of one-half its mass/number of particles. An asymmetric distribution emerges in (a) as the clusters accelerate towards each other. A skewed double-peaked profile is produced in (b). Eye graphics indicate the line of sight. This simulation included 10,000 particles with less than 1% energy error.

4. DISCUSSION

The reader may notice that our simulations are essentially classical, collisionless n -body simulations with few frills and may question the applicability of our results in our study of the dynamical evolution of galaxy clusters. Here we emphasize that our simulations may only hope to demonstrate and establish principles of fairly simple merging scenarios that must be applied with appropriate caution to the more complex processes observed in nature. If we make the assumption that galaxies from which we obtain the radial velocity distribution trace out the mass and radial velocity distributions well, our results will provide a modest estimation of general shapes for which we may look as we

seek to find correlations between our findings and the observed radial velocity profiles of clusters of galaxies.

We have also mentioned the presence of fully equilibrated or relaxed clusters in the initial conditions of our simulation though merging galaxy clusters in nature are unlikely to contain systems which have achieved complete equilibrium. The reader is reminded that the scope of the study is to strictly investigate the merging process, the choice of equilibrium structures in the initial conditions being motivated to control other processes which might affect the resulting radial velocity distributions. We have essentially studied the merging process in its purity, neglecting other dynamical effects. Though our initial conditions are extreme and artificial, we expect that the complex merging systems in nature would not be exempt from exhibiting the signatures presented in our simulations. It is also true that if the notion that large scale structures (such as clusters of galaxies) are formed through the merging of smaller systems is correct, smaller building block systems could have had time to move toward equilibrium allowing for conditions in nature which would bear sufficient resemblance to our contrived scenarios.

In general, we find good agreement between our results and the results included in [9]. By greatly scaling the number of particles, we also observe the very short time frame over which merging processes smooth the radial velocity distribution of the two interacting clusters. Interestingly, though the processes at hand may appear dramatic, none of the radial velocity distributions appear erratic, but rather smooth in character. From our simulation in Figure 2, we also found evidence that a double-peaked radial velocity distribution is effectively erased in under two-thirds the time required to accelerate the cluster to sufficient speeds to produce these spreading distributions. This is again in accordance with [9] who propose that the presence of skew or even "multi-peakedness" in radial velocity distributions of clusters of galaxies could be indicative of an *impending* collision of clusters taking place along the line of sight. Events beyond this stage of impending collision will have effectively erased these features in the radial velocity profile.

Taking skewness of a radial velocity distribution as evidence of a merger has been practiced in [2] where the authors discuss the M86 and M87 "subclumps" of the Virgo Cluster and suggest that a merger is likely taking place between the two subclumps on account of the strongly skewed profile of the dwarf galaxy population in the neighborhood of these galaxies. As of this writing, the authors are not aware of a study in which *multi-peaked* velocity distributions have been taken as evidence of imminent mergers. Though the Virgo Cluster of Galaxies as a whole is irregular and very complex, it might be possible to

apply some of the simple principles discussed above to produce a profile very reminiscent of the velocity profile of the entire Virgo Cluster as it appears in Binggeli's Figure 3 (shaded) in [2]. We assume a cluster of mass 1 with a foreground and a background cluster each with a mass of 0.5. Each cluster is provided with an initial velocity of 0.5 in the direction of the larger cluster and we allow the simulation to run for 9,000 time steps. Figure 4 displays our results for all 10,000 particles in our simulation side-by-side with Binggeli's radial velocity distribution for cluster galaxies. We note remarkable similarities.

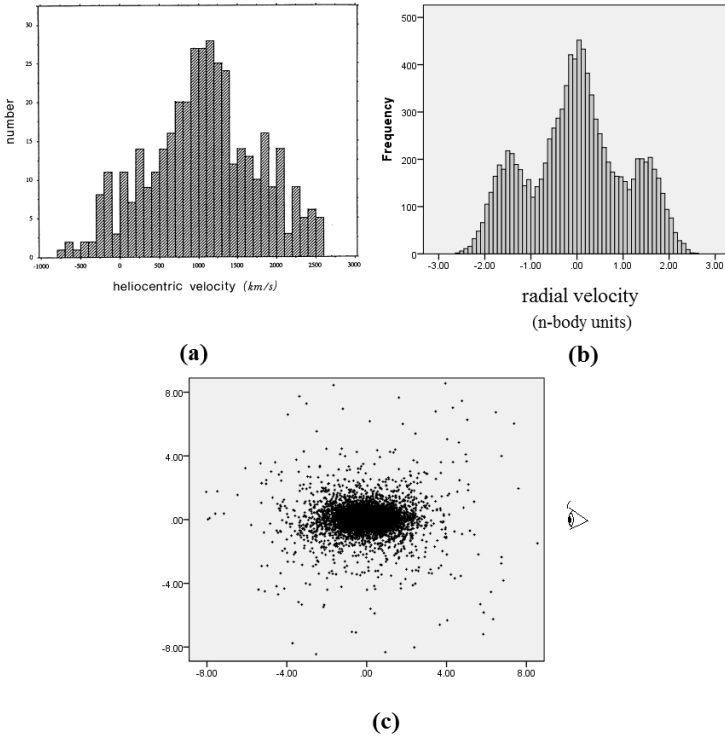


Figure 4: The radial velocity distribution of the entire known membership of the Virgo Cluster appears in (a) (Binggeli, B., Popescu, C. C. and Tammann, G. A. *A&A*, 98, 288, 1993, reproduced with permission © ESO). We compare this distribution with the radial velocity distribution produced from our own simulations in (b) which involved two lower mass clusters merging with a central, higher-mass cluster along the line of sight (see (c)). This simulation contained 10,000 particles and an energy error of 0.17%.

Interpreting this correlation or drawing conclusions is more tricky as a simple merger scenario would be an embarrassing oversimplification of the complexities of the processes taking place within the cluster. In Binggeli's Figure 11 [2], spiral and irregular galaxies also exhibit a triple-peaked radial velocity distribution and also seem to be uniformly scattered over the plane of the cluster. We might postulate that there could exist a foreground and background collection of smaller groups of gravitationally bound galaxies spread over the whole area of the sky subtended by the cluster which are each falling into the cluster's core near M87. The correlation might be painting a picture of a cluster of galaxies which is thus elongated along the line of sight somewhat like the spatial distribution of particles in our simulation. The idea might not be totally without merit. A NED database search of 30 arcminutes around M99 (a galaxy in the Virgo Cluster), for instance, reveals a system of about 10 objects with a central redshift much higher than the mean radial velocity for the entire membership of the Virgo Cluster [7]. If M99's group were *foreground* galaxies, it would not be too difficult to believe that M99 and its associated galaxies are falling away from the observer toward M87 (the galaxy at the center of the cluster) with a significant component of its motion along the line of sight. If enough such galaxy neighborhoods over the entire cluster where falling towards the core from both in front and behind the cluster, we might expect to see the wings of Binggeli's distribution. The authors emphasize, however, the near impossibility of making any definitive statement regarding the processes taking place within the cluster from this type of superficial analysis. This proposal will have to be explored in much greater detail and be supported by x-ray and other applicable evidence. The strong correspondence between our simple simulation and the observed profile is nevertheless presented for the reader's consideration.

A profile which might be read as also having a triple-peaked character exists for the giant galaxies in the Coma Cluster though the radial velocity distribution seems much closer to normally distributed (see Fig 1 in [4]). The appearance of the peaks (the author's choose to point out the appearance of gaps) remains a dynamical mystery. If significant substructure exists in the Coma Cluster, merging scenarios along the line of sight may be worthy of attention.

5. CONCLUSION

During the course of this study, we have investigated mergers between gravitating and collisionless clusters of particles and the resulting radial velocity profiles which would be observed on a collision axis.

For two clusters of equal mass/membership colliding along the line of sight, we find a broadening and an evolution to a double-peaked radial velocity distribution during the moments before collision. This broadening is quickly erased as clusters pass through one another and evolve toward equilibrium. The fact that the appearance of these features is brief could be exploited to identify systems in which subclusters are preparing to merge. The appearance of the double peaks we have observed in our simulations was contingent upon clusters moving at sufficiently high relative speeds. We have also verified that a skewed (and occasionally double-peaked) radial velocity profile is manifest with clusters of unequal mass or membership. In short, we verify the findings of [9] by simulating similar scenarios with many times the number of particles and improved softening techniques. We have further found that more complex scenarios involving more than two gravitating clusters may also produce apparently double-peaked distributions provided the additional cluster is moving perpendicularly to the line of sight. We have shown that additional clusters may serve to aid in accelerating those clusters moving along the line of sight to speeds which might be sufficiently high to produce a more pronounced multi-peaked radial velocity distribution. This latter simulation would also imply that situations which would appear to be simple merger scenarios along the line of sight (as observed and discussed, for example, in [2]) could be more complex than initially perceived.

The foregoing evidence might make it possible to interpret the radial velocity profile for the entire Virgo Cluster in terms of a merger scenario, in which possibly small foreground and background galaxy populations fall into a central and larger subcluster. Although the plausibility of this proposal has not yet been fully investigated and is likely a terrible oversimplification of the system, the correlation between the simulation and the observed radial velocity profiles appears compelling, and, if further investigated, might offer some additional insight into the history and process involved in the formation of the Virgo Cluster of galaxies.

6. ACKNOWLEDGMENTS

We would like Dr. Michael Grady for his continually firing up the supercomputer for our simulation work, Dr. Saïd Bahi for his help with the statistical software used to generate figures in this paper, and Dr. Bryce Christensen for his willingness to review the manuscript and his helpful corrections. A very special thanks is due to the anonymous referees whose feedback greatly improved the quality of this work.

REFERENCES

- [1] Aarseth, S. 2005, *Gravitational N-body Simulations: Tools and Algorithms*. Cambridge University Press, New York.
- [2] Binggeli, B., Popescu, C. C. and Tammann, G. A. 1993, The kinematics of the Virgo Cluster revisited. *Astronomy and Astrophysics Supplementary Series* 98:275-296.
- [3] Binney, J. and Tremaine, S. 2008, *Galactic Dynamics*, 2e. Princeton University Press Princeton, New Jersey.
- [4] Calberg, R. G., Yee, H. K. C., Ellingson, E., Morris, S. L., Abraham, R., Gravel, P., Pritchet, C. J., Smecker-Hane, T., Hartwick, F. D. A., Hesser, J. E., Hutchings, J. B. and J. B. Oke. 1997, The Average Mass profile of Clusters of Galaxies. *Astrophysical Journal* 485: L13-L16.
- [5] Edwards, S., Colless, M., Bridges, J., Carter, D., Mobasher, B. and Poggianti, B. 2002. Substructure in the Coma Cluster: Giants versus Dwarfs. *The Astrophysical Journal* 567: 178-187.
- [6] Hut, P. and Makino, J. 2007. Initial Conditions: Plummer's Model. <http://www.artcompsci.org/kali/vol/plummer/volume11.pdf>
- [7] California Institute of Technology. Nasa/Ipac Extragalactic Database (NED). <http://nedwww.ipac.caltech.edu>
- [8] NVIDIA CUDA Programming Guide: Version 2.0. 2008.
- [9] Schindler and Böhringer. 1993. Simulations of the evolution of galaxy clusters I. The dynamics of galaxies. *Astronomy and Astrophysics* 269: 83-95.
- [10] Schneider, P. 2006, *Extragalactic Astronomy and Cosmology: An Introduction*. Springer, New York.
- [11] Springel, V, White, SDM, Jenkins, A, Frenk, CS, Yoshida, N, Gao, L, Navarro, J, Thacker, R, Croton, D, Helly, J, Peacock, JA, Cole, S, Thomas, P, Couchman, H, Evrard, AE, Colberg, J, Pearce, F. 2005. Simulations of the formation, evolution and clustering of galaxies and quasars *Nature*, 435: 629-636. <<http://hdl.handle.net/2027.42/60629>>

7. APPENDIX

7.1 The Plummer Model

The mass within a given radius r in a Plummer Model is

$$m(r) = r^3(r^2 + 1)^{-3/2}. \quad (7.1)$$

if the mass of the system M and constant scaling factor a are such that $M = a = 1$ [6]. Solving (7.1) for r we obtain

$$r(m) = (m^{-2/3} - 1)^{1/2}. \quad (7.2)$$

If m is uniformly random on the interval $[0,1]$, (7.2) will produce a distribution that will reflect the density profile of the Plummer Model. We populated position space for each Plummer Cluster in the following manner. A random number on the continuous closed interval $[0,1]$ was generated and a corresponding radius was computed through (7.2)

which was multiplied by $r \frac{3\pi}{16}$. The particle was then randomly assigned spherical coordinates ϕ and θ in a manner so that each point on a sphere of radius $r \frac{3\pi}{16}$ would be an equally probable realization.

To populate velocity space, we used the following algorithm from [6] and [1]. Two random variables $\alpha \in [0, 1]$ and $\beta \in [0, 0.1]$ were generated. We define some quantity g as

$$g = \alpha\sqrt{2} \left[1 + \left(r \frac{3\pi}{16} \right)^2 \right]^{-1/4}. \quad (7.3)$$

Random variables α and β are continuously generated until α and β are such that $\beta < \alpha^2(1 - \alpha^2)^{7/2}$. If this condition is satisfied, the velocity

v of the particle at radius $r \frac{3\pi}{16}$ is

$$v = \frac{\alpha \sqrt{\frac{32}{3\pi}}}{\left(1 + \left(r \frac{3\pi}{16} \right)^2 \right)^{1/4}}. \quad (7.4)$$

The velocity was then randomly oriented in three-dimensional space.

7.2 Direct Summation Poisson Solver

If we consider a system of mutually gravitating objects, the effective acceleration experienced by the α th particle is given by

$$\mathbf{a}_\alpha = \sum_{\beta \neq \alpha} G m_\beta \frac{\mathbf{x}_\beta - \mathbf{x}_\alpha}{|\mathbf{x}_\beta - \mathbf{x}_\alpha|^3} \quad (7.5)$$

where G is the gravitational constant, m_β is the mass of the β th particle and \mathbf{x}_i represents the vector position of the i th particle. We note that (7.5) becomes singular for $|\mathbf{x}_\beta - \mathbf{x}_\alpha| \rightarrow 0$. With a large number of particles and a sufficiently long simulation time, we would expect that a force computation will eventually be carried out between two particles with a very small separation, leading to issues with energy conservation within the system. It is common practice to replace (7.5) by the alternative expression

$$\mathbf{a}_\alpha = \sum_{\beta \neq \alpha} G m_\beta S'(|\mathbf{x}_\beta - \mathbf{x}_\alpha|) \frac{\mathbf{x}_\beta - \mathbf{x}_\alpha}{|\mathbf{x}_\beta - \mathbf{x}_\alpha|} \quad (7.6)$$

where $S'(|\mathbf{x}_\beta - \mathbf{x}_\alpha|)$ is the derivative of some function we call a *softening kernel* which will smear out the mass of the β th particle into a volume-occupying cloud so that the gravitational exertion between the particles in question will not diverge with decreasing distance. We used the softening kernel advocated in [3], which is

$$S(|\mathbf{x}_\beta - \mathbf{x}_\alpha|) = -\frac{|\mathbf{x}_\beta - \mathbf{x}_\alpha|^2 + \frac{3}{2}\epsilon^2}{(|\mathbf{x}_\beta - \mathbf{x}_\alpha|^2 + \epsilon^2)^{3/2}} \quad (7.7)$$

where ϵ is a carefully chosen constant called the *softening length*. Differentiating (7.7) and substituting the result into (7.6) we obtain (after routine manipulations) the adjusted algorithm for computing the force used in this study given by

$$\mathbf{a}_\alpha = \sum_{\beta \neq \alpha} G m_\beta \left[\frac{(\mathbf{x}_\beta - \mathbf{x}_\alpha)(5\epsilon^2 + 2|\mathbf{x}_\beta - \mathbf{x}_\alpha|^2)}{2(\epsilon^2 + |\mathbf{x}_\beta - \mathbf{x}_\alpha|^2)^{5/2}} \right]. \quad (7.8)$$

The reader will notice that (7.8) is nonsingular at $|\mathbf{x}_\beta - \mathbf{x}_\alpha| = 0$ (i.e. evaluates to zero) and that the force between any two particles falls off as $\frac{1}{|\mathbf{x}_\beta - \mathbf{x}_\alpha|^2}$ for large values of the quantity $|\mathbf{x}_\beta - \mathbf{x}_\alpha|$ as we would expect for long-distance gravitational interactions.

7.3 *Integrator*

To advance particles in position and velocity space from \mathbf{x} to \mathbf{x}' and \mathbf{v} to \mathbf{v}' , respectively, in a given time step of length h when given acceleration \mathbf{a} (as determined from the Poisson Solver) we used

$$\mathbf{v}_{1/2} = \mathbf{v} + \frac{1}{2}h\mathbf{a}, \quad (7.9)$$

then

$$\mathbf{x}' = \mathbf{x} + h\mathbf{v}_{1/2}, \quad (7.10)$$

and finally

$$\mathbf{v}' = \mathbf{v}_{1/2} + \frac{1}{2}h\mathbf{a}. \quad (7.11)$$

Generation Noner: A Look at Religiously Unaffiliated Attitudes at Salt Lake Community College

Karine Agajanian

University of Utah

Spencer H. Blake

Salt Lake Community College

Abstract

The purpose of this study was to identify the similarities and differences on factors influencing religious non-affiliation between respondents in a highly religious area (i.e., Utah) and those in the nation as a whole. In the early 1990s, around 7% of Americans claimed no religious affiliation, a number that has increased threefold since. According to an extensive study done in 2010, this rise in religious nonaffiliation is due to the increasingly tolerant attitudes of the public towards homosexuality, same-sex marriage, and other social issues as their would-be churches remain intolerant. Students at Salt Lake Community College were surveyed to find whether their reasons for non-affiliation were similar to those in the rest of the country. Results show that social issues such as homosexuality and abortion were the least cited reasons for SLCC students' religiously unaffiliated statuses.

The Research Problem

The purpose of this study is to identify the similarities and differences in factors influencing religious non-affiliation between respondents in Utah and those in the rest of the nation.

Introduction

In 2002, sociologists Michael Hout and Claude Fischer from the University of California, Berkeley, noticed a significant rise in the number of young people who declare they have no religious affiliation. This group is the most rapidly growing religious category in the U.S. today.

In the early 1990s, around 7% of Americans claimed no religious affiliation, a figure that virtually remained unchanged until the next century. New research has shown that number increasing to over 17%, which is 3.5 times higher than any previous cohort. Much of this change is happening in the younger generation; people in their twenties account for 30% of the total nonreligious increase.

In an article in the *LA Times* based on their extensive study, David Campbell and Robert Putnam (2010) pointed out that “the association between religion and politics (and especially religion’s intolerance of homosexuality) was the single strongest factor in this portentous shift.” They go on to say, “Predictions of the demise of religion in America would be premature. More likely is that as growing numbers of young Americans reject religious doctrine that is too political or intolerant for their taste, innovative religious leaders will concoct more palatable offerings.”

Statement of Problem

The purpose of this study is twofold: 1) to identify the factors and the strength of those factors in influencing non-affiliation in students at Salt Lake Community College (SLCC), and 2) to understand whether religious attitudes on political issues in Utah, more specifically in Salt Lake City, are leading residents to shift to the growing non-affiliated status.

Salt Lake City was chosen because Utah has the distinction of being the most religious state in the nation. Does being in a highly religious area create a unique social environment for the non-affiliated? Does living among so many affiliated friends and acquaintances produce a different set of attitudes on being non-affiliated?

Definitions

In this study, the term “noner” means one who is religiously unaffiliated or one who does not identify with any religious denomination.¹ However, this hardly defines the noners completely, as they are a very diverse group in terms of beliefs. Therefore, we employed the options from the religious identification question off of Dr. Karen Franklin’s questionnaire (2000). Each option was different enough to subdivide the noners into four different categories. First, the participants who chose the first option, “I belong to an organized religion,” will be called ‘religious’ and are not included as a noner category. Those who chose the second option, “I believe in God, but don’t belong to an organized religion,” will be called ‘believers.’ Those who chose “I believe in the spiritual, but not religion or God” will be called ‘spirituals.’ Those who chose “I am agnostic” will be ‘agnostic’ and those who chose “I am an atheist” will be ‘atheists.’

Justification of Study

Salt Lake City is unique in that it has a large Mormon population. “Jews and Mormons can be easily recognized through self-identification,” states Campbell and Putnam in their article in the *Los Angeles Times*. “While they are each a much smaller share of the population than Protestants or Catholics, both are highly distinctive traditions that, because of their size, are often neglected in analyses of the American religious environment.” Therefore, understanding whether attitudes within highly religious areas differ becomes pivotal.

Review of the Literature

In 2008, the American Religious Identification Survey found that about 34 million Americans did not have a religious preference, a 138% increase from 1990 to 2008, making it the fastest growing religious category in the nation (Kosmin and Keysar 2008). An extensive study by Campbell and Putnam reported in *American Grace* in 2010 found that about 22% of college freshmen chose “none” as their religious preference in comparison with only 13 percent in 1990. This shift is predominantly happening in the younger generation, those who have come of age since 1990.

¹ The term “noner” has been used extensively in literature dealing with growing non-affiliation in recent decades.

As mentioned above, many studies across the nation have suggested that this rise in “noners” has come about because of the increasingly conservative role religion is playing in politics (Hout and Fischer 2002; Chaves and Cann 1992; Campbell and Monson 2005; Wuthnow 1999). Specifically, the younger generation’s support of homosexuality and gay marriage in contrast to their parents is thought to be one of the strongest reasons for this religious shift. The 2004 presidential elections were an excellent example of the politicalization of religion. President Bush was said to have “pulled the religion card” in attempts to persuade four million evangelicals to the polls; this proved most effective in predominantly evangelical states with a gay marriage ban on the ballot, illustrating a link between one’s politics and religion (Campbell and Monson 2005). As stated by Campbell and Putnam in the *Los Angeles Times*,

“Just as this generation moved to the left on most social issues—above all, homosexuality—many prominent religious leaders moved to the right, using the issue of same-sex marriage to mobilize electoral support for conservative Republicans ... Increasingly, young people saw religion as intolerant, hypocritical, judgmental, and homophobic. If being religious entailed political conservatism, they concluded, religion was not for them.”

Despite a wide range of research based on the nation as a whole, very little attention has been paid to religious non-affiliation in highly religious communities. It is unknown whether those who do not identify with a religion in these areas follow the national trends and whether their reasons for non-affiliation are the same. Hadden and Long (1983) stated, “One important consequence of secularization...is the erosion of the authority of religious institutions to confer legitimacy.” While the rest of the nation may be experiencing this erosion and moving away from religion, highly religious areas may not be affected. In their recent work, *American Grace*, Campbell and Putnam found that Utah scored higher in religiosity than any other state in the nation. This was measured by such questions as “How frequently do you attend religious service?” “How frequently do you pray outside of religious services?” and so on. It is also the only state with such high uniformity of religion. The Mormon faith, also known as the Church of Jesus Christ of Latter-Day Saints (LDS), is the most prominent religion in Utah. Those who are religiously unaffiliated and who live in such a highly religious area such as Utah may not have the same reasons for non-affiliation as the rest of the country.

Research Procedures

Sampling Procedures

The population for this study was students enrolled at SLCC in Spring Semester 2011. Demographic information, such as age, race, sex, religious affiliation growing up, religious affiliation currently, political affiliation, political outlook, parents' education levels, and frequency of religious attendance within the past year, was collected. This study used survey research to collect data. The questionnaire (Appendix) was a collaborative effort with another research group, duplicating some demographic questions from Karen Franklin's (2000) study. Nineteen questions were added to the Franklin study. The questions used Likert scales consisting of (1) strongly agree, (2) somewhat agree, (3) somewhat disagree, and (4) strongly disagree, or favor/oppose questions instead of agree/disagree where appropriate.

The questionnaires were given out across various SLCC general education classes anonymously for students who would voluntarily like to participate, finishing them at home and returning them the next class meeting.

Data Analysis

For analyzing the collected data, the Statistical Package for Social Sciences (SPSS) 14.0 and Predictive Analytic SoftWare (PASW) 18.0 were used. Originally, questionnaires were distributed to about 500 participants; 420 participants completed and returned them. Using SPSS, frequencies and cross-tabulations were employed to generate accurate numbers, and one-way ANOVA tests were used to compare the religious with the nonreligious as well as the four different subdivisions of noners for statistical significance between groups.

Demographics

The sample included 420 participants, with 43 percent male and 57 percent female and a mean age of 24 years, ranging from 18 to 59 years. About 77% identified as White, 8% as Asian, 7% as Latino, 2% Black, and 1% Native American; an additional 4% identified as "mixed" or "other." Out of the 57.7% that fell under the 'religious' category, 77.5% identified as LDS.

Dividing the nonreligious, 178 participants out of 420 were categorized as "noners" (42.3%). Forty-three percent were male and 57% were female with a mean age of 23 ranging from 18 to 50. Regarding

ethnicity, 77% were White, nine% were Latino, nine% were “mixed” or “other”, three% were Asian, two% were Black, and two% were Native American.

Table 1. Political Affiliation by Group (%)				
	Repub.	Dem.	Indep.	Other
Religious	49.5	11.1	23.1	16.2
Noner	14.7	30	28.2	27.1
Believer	19.2	31.5	23.3	26
Spiritual	17.2	27.5	38	17.2
Agnostic	12	36	20	32
Atheist	4.1	29	29	37.5

Separating the noners shows that out of the 178 participants, 43% were “believers,” 17% were “spiritualists,” 15% were “agnostic,” 15% were “atheist,” and 11% were “other.” Tables 1 and 2 show political affiliation and political/social outlook divided first by religious and noner, and then subdividing the noners into believer, spiritual, agnostic, and atheist. The noner category is the average combined scores of believers, spirituals, agnostics, and atheists. Exact numbers are shown in Figure 1.

Table 2. Political/Social Outlook (%)			
	Conservative	Middle of the road	Liberal
Religious	51.6	30.5	17.8
Noner	18.5	38.6	42.8
Believer	27.3	38.4	34.2
Spiritual	10.7	25	53.5
Agnostic	16	32	48
Atheist	12.5	20.8	62.5

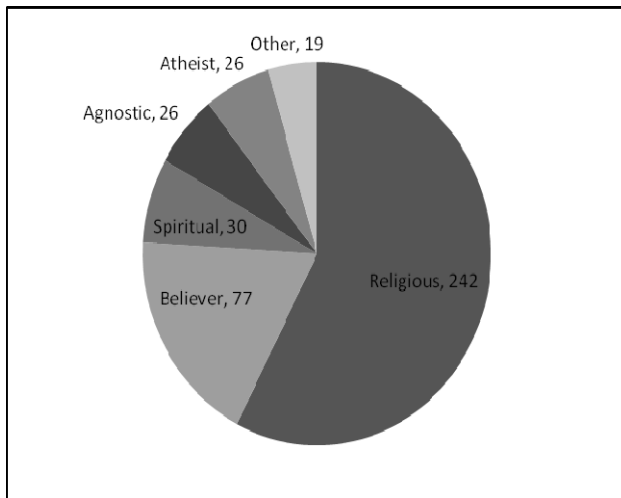


Figure 1. Which of the following best describes your current religion?

Results

A staggering 42.3% of the sample fit into the “noner” category (178 out of 420), much higher than Campbell and Putnam’s 22% of college freshmen. When asked whether they would favor electing religious leaders, 75.1% of the religious said they would favor it while only 23.5% of the noners would favor it. When asked whether deeply religious leaders made better decisions than nonreligious leaders, 65.7% of the religious agreed that they do make better decisions, while only 18.2% of the noners agreed with the statement. About 60% of the religious and 78% of the noners agreed that religious leaders should not influence voting. Fifty-three percent of the religious and 75% of the noners agreed that religious leaders should not influence government decisions. Results are shown in percentages in Table 3.

Finally, participants were asked to what extent the following factors influenced their religious affiliations: parental influence, peer influence, group identification, homosexuality, abortion, women’s rights, race relations, philosophical views, scientific views, and lifestyle conflicts. The results comparing the religious and the noners are shown in Table 4 and Figure 2. An ANOVA test showed a statistically significant difference between religious and noner responders in the peer influence and group identification options ($p < 0.005$), with parental influence and philosophical views coming very close to statistical sig-

Table 3. Questions regarding religious leaders					
Relig.	Noner	Believer	Spirit.	Agnos.	Atheist
16. Favor electing religious leaders					
75.1	23.5	30.4	7.1	19.2	15.3
17. Agree that deeply religious leaders make better decisions than non-religious leaders					
65.7	18.2	19.7	10.3	20	11.5
18. Agree that religious leaders shouldn't try to influence voting					
59.7	78.4	75	86.2	88.4	73.1
19. Agree that religious leaders should not influence government decisions					
52.8	74.4	68.4	89.6	84.6	73

nificance ($p=0.056$). Differences between the two groups in regards to the social issues such as homosexuality and abortion showed no statistical significance.

The results between the subdivisions of noners are shown in Table 5 and Figure 3. An interesting difference was observed between believers and spirituals. There was strong statistical significance with *parental influence*, *group identification*, *philosophical views*, and *scientific views* ($p < 0.05$); however, as the samples for the spiritual, agnostic, and atheist groups were so small compared with the believers,

Table 4. Influences on Religious Affiliation between Noners and Religious		
Influences	Noners	Religious
Parental	46.6	68.8
Peer	27.8	48.4
Group ID	18.6	44.1
Homosexuality	17.8	17
Abortion	15.1	20.6
Women	24.9	28.8
Race relations	20.8	24.4
Philosophy	48.9	46
Science	46.8	31.5
Lifestyle	49.7	42.7

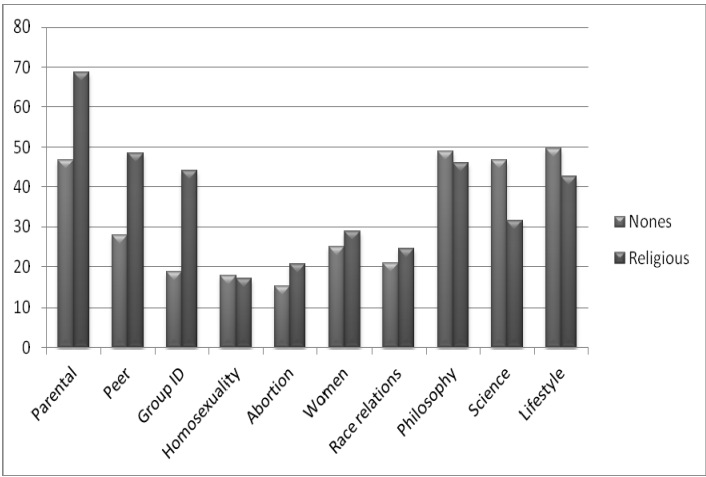


Figure 2. Influences on religious affiliation between noners and religious

these differences may fade with larger samples. Also, atheists were the only group to score significantly higher on *homosexuality* and *abortion*. This suggests that atheists may be the group who most identify with and follow the national trend; however, they were still well below the 50% mark on both homosexuality and abortion.

Table 5. Influence on Religious Affiliation within Noner Category				
	Believers	Spiritual	Agnostics	Atheists
Parental	52	35.7	38.5	38.5
Peer	28.1	14.2	34.6	23.1
Group ID	19.1	0	23.1	19.2
Homosexuality	13.7	10.7	15.4	38.4
Abortion	11.1	11.1	11.5	30.7
Women	20.5	29.6	19.2	38.4
Race relations	21.9	14.8	15.3	30.7
Philosophy	33.8	70.3	61.5	61.5
Science	32.9	59.2	53.8	69.2
Lifestyle	53.4	44.4	32.3	53.9

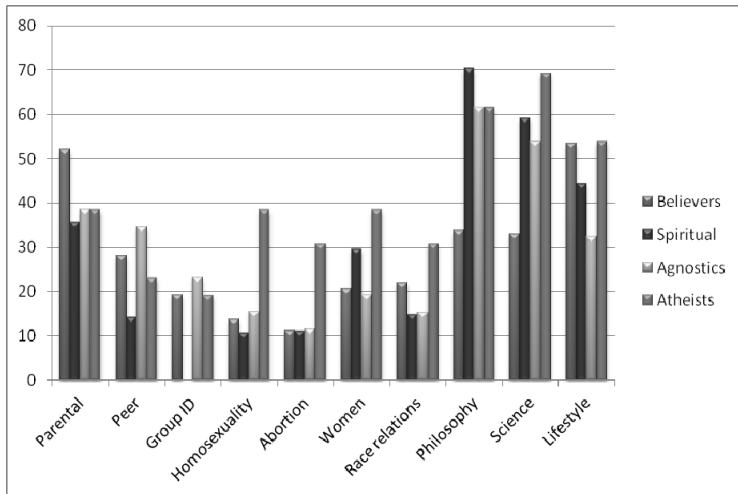


Figure 3. Influence on religious affiliation within noner category

Limitations

No empirical investigation of social phenomena is free from limitations or qualifications. The following is a short rehearsal of some limiting factors regarding this study. The focus of this study was to ascertain attitudinal differences. We gave a list of possible influences for the participants' attitudes regarding religion, but did not track the direction of that influence. For instance, a believer could have said that *parental influence* was a significant contributor to their noner status; whether his or her parents influenced them into religion or out of religion cannot be determined from our questionnaire. We believe that the religious and noners may have had far greater differences between them had we looked into the direction of their influences. Another limitation came from answers respondents gave. In a few cases it became clear that some respondents did not understand the differences between agnostic and atheist. This challenged the coding of a few responses. As this was only a correlational study, cause could not be conclusively discerned.

Summary and Conclusions

Being a highly religious area, Salt Lake City gave a unique set of variables to further understand the increasing trend of non-affiliated status in the U.S. Two are noteworthy. First, the national trend of relig-

iously unaffiliated 20-somethings hovers just under 30%. Our results showed 42%. It is possible that a highly religious area like Salt Lake City is creating a push on those “sitting on the fence” (with regards to religious affiliation) to take sides. As religious areas tend to measure one’s identity with one’s religious affiliation, those who are undecided may feel pressured to choose. On the other hand, these results may only be illustrating the continuing rise in rates of the non-affiliated.

Secondly, this study expected to find attitudinal and influential differences between noners in the nation and noners in Salt Lake City. Our results show that the reasons for non-affiliation are not the same as those outside this region. What was surprising was the direction that this difference took.

Had Utah been following national trends, we would have expected to see higher scores on sociopolitical factors such as *homosexuality* and *abortion*. Instead, the factors that noners chose most were *parental influence*, *philosophical views*, *scientific views*, and *lifestyle conflicts*. In fact, *homosexuality* and *abortion* were the least chosen out of both religious and nonreligious participants. This aspect seems to be the opposite of the results found when looking at the nation as a whole.

If the unaffiliated in Utah are not finding their politics to be at odds with religion, this could mean that the effects of secularization have not eroded religious authority in Utah to the same degree as they have elsewhere in the nation. It could suggest that religion is an unimportant factor for attitudes in Utah, if not for the blatant shift in politics and sociopolitical outlook when looking from the religious to the noners.

It suggests that the institution of religion in Salt Lake City is still able to compete with sociopolitical issues, at least on *homosexuality* and *abortion*. Religion in Utah, then, is still providing acceptable answers of *why* and *what ought to be* even among the unaffiliated.

This can come about for two reasons: one, the message that is being delivered is consistent with the receivers’ general value orientations, and two, to some degree it is consistent with their own interest.

Whether the values (instilled in the non-affiliated) are a holdover from growing up in a religious home or are just consistent with the individual’s own interest, both give a *voice* that is consistent within the non-affiliated psyche.

Acknowledgments

We would like to express gratitude to both the Salt Lake Community College professors who allowed us to come into their classes and to the students who participated in the study. Additionally, special

thanks go to CoCo Bronson James, Chad Morrow, and Michael Cuppello for their assistance with SPSS data entry.

Bibliography

Campbell, David E. and Monson, J. Quin. The Case of Bush's Re-election: Did Gay Marriage Do It? <http://www.nd.edu/~dcampbe4/bush.pdf>. Unpublished.

Campbell, David E. and Putnam, Robert D. *American Grace: How Religion Divides and Unites Us*. (New York, NY: Simon & Schuster, 2010).

Campbell, David E. and Putnam, Robert D. "Young Americans Leaving Churches That Push Political Agenda". *Los Angeles Times*. Oct. 31, 2010.

Chaves, Mark and Cann, David E. "Regulation, Pluralism, and Religious Market Structure: Explaining Religion's Vitality" *Rationality and Society*, Vol. 4, (1992), 272-290.

Franklin, Karen. "Antigay Behaviors Among Young Adults: Prevalence, Patterns, and Motivators in a Noncriminal Population." *Journal of Interpersonal Violence*, Vol. 15, No. 4, (2000), 339-362.

Hadden, Jeffrey K. and Long, Theodore E. *Religion and Religiosity in America*. (New York: Crossroad Publishing Company, 1983).

Hout, Michael and Fischer, Claude S. "Why More Americans Have No Religious Preference: Politics and Generations," *American Sociological Review*, 67, (2002), 165-190.

Kosmin, Barry A. and Keysar, Ariela. *American Religious Identification Survey: Summary Report*. (Hartford: Trinity College, 2008).

Wuthnow, Robert. "Mobilizing Civic Engagement: The Changing Impact of Religious Involvement." In *Civic Engagement in American Democracy*, edited by T. Skocpol and M.P. Fiorina. 331-63. Washington, DC: Brookings Institution Press (1999).

Appendix

Questionnaire

1. Which of the following best describes your current religion?
 - a) I belong to an organized religion
What religion? _____
 - b) I believe in God, but don't belong to an organized religion.
 - c) I believe in the spiritual, but not religion or God.
 - d) I am an agnostic.
 - e) I am an atheist
 - f) Other (please explain) _____
2. To what degree is your affiliation and/or church attendance, or lack thereof, influenced by the following factors?

	Strongly Disagree	Somewhat Disagree	Somewhat Agree	Strongly Agree	
	1	2	3	4	
a. Parental influence		1	2	3	4
b. Peer influence		1	2	3	4
c. Group identification		1	2	3	4
d. Homosexuality		1	2	3	4
e. Abortion		1	2	3	4
f. Women's rights		1	2	3	4
g. Race relations		1	2	3	4
h. Philosophical views		1	2	3	4
i. Scientific views		1	2	3	4
j. Lifestyle conflicts		1	2	3	4
k. Did a single episode or experience affect this?		1	2	3	4

2. What age were you when you decided on your current religious affiliation?
3. Generally speaking, do you usually think of yourself as Republican, Democrat, Independent, or something else?

4. Thinking politically and socially, how would you describe your own general outlook—as being very conservative, moderately conservative, middle-of-the-road, moderately liberal, or very liberal?

5. How much are you in favor of electing deeply religious leaders (of any faith)?

Strongly Oppose	Somewhat Oppose	Somewhat Favor	Strongly Favor
1	2	3	4

6. How much do you agree that deeply religious leaders make better decisions than non-religious leaders?

Strongly Disagree	Somewhat Disagree	Somewhat Agree	Strongly Agree
1	2	3	4

7. How much do you agree that religious leaders should not try to influence how people vote?

Strongly Disagree	Somewhat Disagree	Somewhat Agree	Strongly Agree
1	2	3	4

8. How much do you agree that religious leaders should not try to influence government decisions?

Strongly Disagree	Somewhat Disagree	Somewhat Agree	Strongly Agree
1	2	3	4

A Second Soul: Exploring Contexts of Language Acquisition and Personality

Aaron K. Combs

Dixie State College

Abstract

The effect of language on personality merits consideration. Studies involving language and personality focus primarily on personality traits that affect language acquisition. In this a paper, I present research on bilingual personalities, especially on the ways in which language serves as a contextual variable that can influence personality. As bilingual speakers shift from one language to another, there may also be a shift in their personalities. In addition, I examine current theories relating to observed shifts in personality when bilinguals switch languages. Finally, I suggest sites for future research that would include investigations into whether or not the shift in bilingual personalities is affected in part by principles of operant and classical conditioning during acquisition.

Charlemagne said, "To have another language is to possess a second soul." In this quote, Charlemagne hints at an important phenomenon that can be seen in people who acquire a second language. They may also begin to develop another identity and personality, a "second soul." In a survey involving 1,039 bilingual individuals, 65% of the participants felt this to be true and responded affirmatively to the question "Do you feel like a different person sometimes when you use your different languages?" (Pavlenko, 2006). This effect of language on personality merits consideration because it suggests that the context in which a language is learned may produce changes in an individual's personality.

Literature Review

Personality is an important theme in psychology. Personality refers to a person's characteristic patterns of behaviors, emotions, and cognition. The majority of studies involving language and personality have focused on the personality traits that affect language acquisition, but relatively few studies examine the effect that learning a new language can have on the individual's personality in the new language. One reason this might be is that personality has traditionally been viewed as relatively stable and unchanging (Allport, 1937). More recent research has shown that the view of personality as invariable is not completely accurate. An individual's behavior can change considerably depending on the situation (Shoda et al. 1989, 1993a, 1993b, 1994). For example, an individual who may be quiet and reserved around his or her family might also be loud and talkative in a different situation, such as when around friends.

Language can be viewed as one of the many important contextual variables affecting personality. Language can influence aspects considered to be part of personality, such as actions, thoughts, and feelings. Shifts in language can also reflect shifts in actions, thoughts, and feelings.

Shifting one's language can shift an individual's pattern of thought. Evidence of this is seen in cultural frame switching. Cultural frame switching is a process in which bilingual and bicultural individuals shift attributions, values, self-concept, and cultural identification when primed with culturally relevant stimuli (Benet-Martinez et al. 2002). For example, when primed with American icons, English-Chinese bilinguals make internal attribution more frequently, which is characteristic of Western culture, and when primed with Chinese icons, English-Chinese bilinguals make external attributions more frequently, which is characteristic of Asian culture. In short, language can be

viewed as a culturally relevant stimulus. When bilinguals shift from one language to another, they often experience a change in decision-making patterns, dogmatism, and self construal (Earl, 1969; Hong et al. 2001). Memories and learning are also affected by language. Memories and learning obtained in one language are more accessible in the language in which they were learned (Marian and Neisser, 2000). This increased accessibility to knowledge and learning in one language as opposed to an alternative language can in turn influence personality, allowing one to draw on different memories depending on the language being spoken.

As bilinguals shift from one language to another, there is also evidence of shifts in emotional expression. Bond and Lai (1987) showed that Hong Kong bilinguals could express their ideas on embarrassing topics at greater length while speaking in their second language. This may be due to varying amounts of emotional restraint seen in bilinguals, depending on which language they are speaking. For example, if an individual learns about bodily maturation or sexual functions in English, then that individual might be more comfortable discussing the topic in a language that is not associated with the emotion or embarrassment that is connected with the topic in English. An individual may be able to discuss the topic more comfortably in another language. Also studies using physiological measures demonstrated that emotionally laden words, such as reprimands, evoke a stronger physiological response when presented in a bilingual's native language, as opposed to when it is presented in the second language, showing that not only emotional expression but perhaps also emotional experience is affected by the language that a bilingual is speaking (Harris et al., 2003).

Other studies indicate that even changing the language in which a personality test is given can be enough of a contextual variable to produce different results depending on the language in which the test is given. Ervin (1964) demonstrated this by administering a Thematic Apperception Test (TAT) of the same picture in French and English to English-French bilinguals. When the TAT was administered in English, achievement themes were more common in female bilinguals. When the test was administered in French to the same individuals, verbal aggression against same age peers, autonomy, or withdrawal from others were more common. In short, the same individuals exhibited contrasting themes while taking the TAT depending on the language in which it was administered. Also, using the Big Five Inventory personality test, researchers found that Spanish bilinguals responding in English responded in a more traditionally perceived American manner and were reported as being more extroverted, agreeable, conscientious, and less neurotic than when the inventory was given in Spanish (Ramirez-

Esparza et al. 2006). This is interesting because it shows that the language in which a personality test is given could influence factors associated with the participant's personality while taking the test.

These studies present evidence that shifting one's language can also result in a shift in one's personality. Studies on language and personality are not isolated to self-report and self-perception studies. Shifts of personality in bilinguals are also observable by others. Pierson and Bond (1982) used standardized interviews with Hong Kong bilinguals and found differences in nonverbal behavior and self-perception of the participant. While speaking Chinese, bilinguals spoke faster, had fewer pauses, increased gazing, and had greater redundancy than when they spoke in English. While speaking English they talked less, smiled more, and had more torso shifts. These nonverbal differences offer overt signals that there has been a shift in the personality of the speaker. As part of another study, researchers videotaped bilinguals while speaking Chinese and while speaking English. Two judges then independently analyzed personality aspects in the participants. When speaking English, bilinguals were rated as higher in extroversion and openness to experience. When speaking Chinese, the same individuals were rated lower on restraint (Chen, 2007). This demonstrates that bilinguals appear to shift personalities when they switch languages.

Theoretical Causes of Language's Effect on Personality

These findings suggest that the phenomenon of language's influence on personality is real, but what causes it? Most studies have focused on cognitive factors such as cultural frame switching (CFS) (Hong et al. 1997, 2000) and communication accommodation theory. These theories explain language's effect on personality by the way language primes bilinguals to identify with a culture or facilitate interpersonal communication and harmony. The theories focus on the individual's expectations and beliefs of the culture associated with the language. In other words, if a bilingual perceives the American culture as being extroverted and assertive, then as that bilingual speaks English he or she will identify with the American culture and speak and act more extroverted and assertive. These theories may explain and help predict language's influence on personality; however, some findings have shown that a bilingual's personality will contradict his or her personal perception of the culture to which the language is associated. For example, Chen (2007) hypothesized that bilinguals speaking Chinese would identify with and be primed towards a personality that would fit their personal perception of the culture associated with Chinese. For example, if they perceived the culture associated with Chinese as being

more reserved, then while speaking Chinese they would act more reserved. Chen found, however, that judges rated bilinguals speaking Chinese as having more traits associated with American culture, such as extroversion, rather than Chinese culture. This finding provides evidence that factors other than CFS or cultural accommodation might be influencing bilinguals' personalities.

How a bilingual develops a second personality while speaking his or her second language may be better understood by considering an individual's personal experience while learning the second language. For example, an individual who is introverted and speaking English could very well perceive his or her own culture as being high in traits such as extroversion and view the need to have that trait, but because of personal experience that person may struggle with behaviors associated with extroversion. By focusing on a person's personal experience in learning personality traits within the context of language, researchers can view personality development in an alternate way. An extrovert could be viewed as a person who has been highly reinforced for seeking social interaction, and those interactions could also have become associated with positive feelings. Thus, the extrovert is more likely to perform social behaviors and is more likely to feel positively towards social interactions. On the other hand, one could reason that a person who is introverted could have been punished for engaging in social interaction, and his or her social contact with others would have become associated with negative feelings.

The context of language can serve as a cue that evokes learned personality traits and elicits emotions that are associated with the context. As the language context is changed, the set of learned personality traits and conditioned emotions associated with that context may also be changed. In other words, if an individual has learned traits and emotions associated with *introversion* in the language context of English, and the same individual has learned traits and emotions associated with *extroversion* in the context of Spanish, then while speaking English, that individual will be primed towards introversion, and while speaking Spanish the individual will be primed towards extroversion.

Future Research

Future research into the relationship between language and personality could provide understanding about the importance of context in learning a language. Learning a language may have implications beyond the acquisition of language skills. Future research may, for example, examine students who learn a language in a study-abroad program. Researchers may benefit from monitoring personality changes,

both self-expressed and observed. For example, do the students spend a significant portion of the day in the classroom? Do they often speak the language in social situations with native speakers? In what context do they learn the language? Do they perceive a change in their own personalities?

One particularly rich study may include returned missionaries from the Church of Jesus Christ of Latter-Day Saints (LDS). Some missionaries for the LDS Church learn a second language. There has been significant research into the effectiveness of the language acquisition of these missionaries who speak the acquired language for approximately two years (Paul, 2009; Yamamoto, 2001; Parker, 2001; Fowler, 1991). There have also been studies that investigate the relationship between culture and language in the returned missionaries, specifically the need for missionaries to be more culturally sensitive as they speak a learned language (Moon, 1981; Hafen, 1987); however, there is a need for studies into the possible effect of the new language on the personality of the speaker. Returning missionaries often comment that they feel a personality difference when they speak the acquired language (Yamamoto, 2001). Most of the studies regarding language surround missionaries who serve in China, Japan, and Korea. Although a majority of returned missionaries learn Spanish, there are few studies that center on Spanish-speaking missionaries. Spanish-speaking returned missionaries may prove especially helpful in identifying connections between language and personality because they have many opportunities to switch back and forth between languages in various contexts after their missionary service is completed.

In this paper, I have presented studies that indicate a possible relationship between language acquisition and personality. This is important because it suggests that the context in which a language is learned may produce changes in an individual's personality. Future research could study these personality changes in returned missionaries from the LDS Church. I reason that because of the highly social context in which LDS missionaries learn a language, the missionaries may acquire highly social personality traits that they do not possess when they speak in their native language. Research could identify whether those highly social personality skills remain with the missionary even when the context changes such as when an introverted English-speaking missionary learns Spanish as an LDS missionary in a highly social context. When that young person completes the missionary assignment, it is possible that when he or she speaks Spanish again, the highly extroverted social skills may be present. In sum, the speakers may exhibit those same extroverted social skills that they acquired as they were learning the language, even when they are no longer in the same social context. The

context in which missionaries originally learned the language may influence their behavior each time they speak the acquired language. Because of the highly social context in which they learn a language, returned LDS missionaries may offer insights into the possible effects of language on personality. Results could lead to future implications about language and personality in general.

Simply realizing that individuals can have access to another personality, or “second soul,” is intriguing. Language is an important context for personality and can be compelling to both researchers and individuals interested in learning another language. Language studies should place more emphasis on *how* a language is learned and not simply on *how fast* it is learned.

References

Allport, G.W. (1937). *Personality: A psychological interpretation*. New York: Holt, Rinehart & Winston.

Benet-Martinez, V., Leu, J., Lee, F., and Morris, M. W. (2002). Negotiating biculturalism: Cultural frame switching in biculturals with oppositional versus compatible cultural identities. *Journal of Cross-Cultural Psychology*, 33, 494-516.

Bond, M.H., and Lai, T.M. (1987). Embarrassment and code switching into a second language. *Journal of Social Psychology*, 13, 169-185

Chen, S.X. (2007). *Two languages, two personalities? Examining language effects on personality in the bilingual context*. [Unpublished doctoral dissertation]. Hong Kong: The Chinese University of Hong Kong.

Earl, M. (1969). A cross-cultural and cross language comparison of dogmatism scores. *Journal of Social Psychology*, 4, 1-20

Ervin, S.M. (1964). Language and TAT content in bilinguals. *Journal of Abnormal and Social Psychology*, 68, 500-507.

Fowler, Elizabeth M. (1991, June 11). Careers: Managers lack fluency in languages. *New York Times* (Late Edition).

Hafen, Jeffrey Kent (1987). Latin American cross-cultural education for missionaries: Latter-day Saint Missionary Training Center, a case

study, Provo, Utah, 1987. Ph.D. dissertation, Brigham Young University, Utah.

Harris, C.L., Aycicegi, A., Cleason, J.B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Applied Psycholinguistics*, 24, 561-579

Hong, Y.Y., Chiu, C.Y., and Kung, T.M. (1997). Bringing culture out in front: Effects of cultural meaning system activation on social cognition. In K. Leung, Y. Kashima, U. Kim, and S. Yamaguchi (Eds.), *Progress in Asian Social Psychology* (Vol. 1, pp. 135-146). Singapore: Wiley.

Hong, Y.Y., Ip, G., Chiu, C.Y., Morris, M.W., and Menton, T. (2001). Cultural identity and dynamic construction of the self: Collective duties and individual rights in Chinese and American cultures. *Social Cognition*, 19, 251-268

Hong, Y.Y., Morris, M.W., Chiu, C.Y., and Benet-Martinez, V. (2000). Multicultural minds: A dynamic constructivist approach to culture and cognition. *American Psychologist*, 55, 709-720.

Marian, V. and Neisser, U. (2000). Language-dependent recall of autobiographical memories. *Journal of Experimental Psychology: General* 129(3), 361-68

Moon, Tae-Im (1981). Shifts and constancies in Rorschach responses as a function of culture and language. Ph.D. dissertation, Brigham Young University, Utah.

Parker, Tyrone Frederick (2001). The perceptions of traditional and nontraditional students toward the level I Spanish textbook. Ed.D. dissertation, West Virginia University, West Virginia.

Paul, M. (2209) Chinese foreign language attrition: Investigating aspect marker usage. Ph.D. dissertation, The University of Arizona, Arizona.

Pavlenko, A. ed. (2006), *Bilingual Minds: Emotional Experience, Expression and Representation*, Clevedon, UK: Multilingual Matters.

Pierson, H.D., and Bond, M.H. (1982). How do Chinese bilinguals respond to variations of interviewer language and ethnicity? *Journal of Language and Social Psychology*, 1, 123-139

Ramirez-Esparza, N. Gosling, S.D., Benet-Martinez, V., Potter, J.P., and Pennebaker, J.W. (2006). Do bilinguals have two personalities? A special case of cultural frame switching. *Journal of Research in Personality*, 40, 99-120

Shoda, Y., Mischel, W., and Wright, J.C. (1989). Intuition interactionism in person perception: Effects of situation-behavior relations on dispositional judgments. *Journal of Personality and Social Psychology*, 56, 41-53

Shoda, Y., Mischel, W., and Wright, J.C. (1993a). Links between personality judgments and contextualized behavior patterns: Situation-behavior profiles of personality prototypes. *Social Cognition*, 4, 399-429.

Shoda, Y., Mischel, W., and Wright, J.C. (1993b). The role of situational demands and cognitive competencies in behavior organization and personality coherence. *Journal of Personality and Social Psychology*, 65, 1023-1035

Shoda, Y., Mischel, W., and Wright, J.C. (1994). Intra-individual stability in the organization and patterning of behavior: Incorporating psychological situations into the idiographic analysis of personality. *Journal of Personality and Social Psychology*, 67, 674-687.

Yamamoto, C. (2001) A comparative discourse analysis of adult Japanese learners and native Japanese speakers. Ph.D. dissertation, Brigham Young University, Utah.

Learning About Peace and Justice Work

Josh Gold

Salt Lake Community College

ABSTRACT

Extensive interviews were conducted during 2008-2009 for the purpose of designing and planning a Peace and Justice Studies (PJS) Curriculum for the School of Humanities and Social Sciences at Salt Lake Community College. Here, I present some of the insights gleaned from professionals whose work in public and civil society organizations in the Salt Lake Valley deliberately promotes peace and justice locally, nationally, and globally. As a result of learning about peace and justice work in the community from them, I propose that such civil society links should help people across the state of Utah address the gender, racial, ethnic, class, and ideological divides that cause conflict and costly divisions in our community. The need for a PJS program to work with students in the development of a cost accounting—of waging peace vs. war in the 21st century—became apparent and, as a result, 2 service-learning Peace and Conflict courses (1 domestic and 1 international) were developed as the anchor of the curriculum. In addition, the sabbatical time and the 77 subjects interviewed helped me to construct a model of how violence operates, which I call “A Political Economy of Direct Violence.” It became another way to teach for peace and justice.

Introduction¹

For the 2008-09 academic year, the Salt Lake Community College (SLCC) administration awarded me a sabbatical to research and write a Peace and Justice Studies (PJS) curriculum for the School of Humanities and Social Sciences. The plan for collecting ideas, information, and insights about doing peace and justice work in Salt Lake City (SLC) included conducting formal interviews with 77 public and civil society workers in SLC, virtually all of whose labors promote peace and justice. The list of interview questions increased across the seven and a half months of interviews, and some interview subjects were queried multiple times. As a consequence, the interviews grew in volume and purpose across the period. Nonetheless, one question that ended each interview required subjects to provide the contact names and information for others in the community they believed should be interviewed on the subject of peace and justice work. This snowballing technique led my queries far and wide in the Salt Lake Valley (SLV), including sitting and talking with the co-chairs of the Utah Minutemen.²

The interviews allowed me to collect information, statistics, ideas, and passionate commitment about peace and justice [studies] work from the following fields human endeavors, among others: domestic violence (DV) matters; refugee and immigrant affairs; Indian affairs; aging and elderly affairs; Utah's food and hunger matters; Latino affairs; Utah's community and cultural affairs; gay, lesbian, bisexual, and transgendered (GLBT) concerns; U.S. civil rights issues; diversity and multicultural affairs; U.S. veterans affairs; U.S. and international humanitarian and human rights issues; service-learning affairs; alternative high school and English as a Second Language (ESL) affairs; matters of international war and peace; jobs with justice and organized labor activities in Utah; SLCC student affairs; Utah Minutemen affairs.

¹ This paper is dedicated to Salt Lake City's finest Peace and Justice activist, Dayne Goodwin. The Valley is lucky to have a man so fundamentally dedicated to stopping war and the social, political, and economic madness it requires. This author and amateur activist would like to thank him for the training and guidance across much of an academic year (2008-2009). Shout-outs to Drs. Stan Taylor, LCC, and Roger Durham, Aquinas College, MI, are also in order. Thanks, brothers, for your inspiration, and your dedication to peace and justice. Finally, SLCC must be thanked for providing this author the school year to do the work herein specified.

² Any reader desirous of the facts, figures, names, dates, interview questions, and actual transcriptions of interview subjects must contact the author: josh.gold@slcc.edu.

All of the formal interviews were undertaken to allow me to learn about the policies, standard operating procedures (SOPs), and costs of purposeful and organized efforts by state and local civil society organizations, as they have been deliberately engaged in building social, cultural, historical, and civic ties in the many communities of Utah. Consequently, the heart of the PJS—the two Peace and Conflict courses (domestic and international)—requires that students volunteer with local civil society organizations working on issues of identified need in the community and beyond. Service-learning courses are one way the SLCC PJS could help students recognize the need for active and direct non-violent means of addressing conflict. This is because such a pedagogy will focus student attention on the costly macro-economic and political disconnections in the larger world (i.e., of the costs and benefits of waging peace vs. war) and provide them with the tools to work toward overcoming them.

Such service-learning placements should help students see how people across the state of Utah address the gender, racial, ethnic, class, and ideological divides that cause conflicts in every community. Moreover, when institutions of the public sector and civil society are vibrant, operating extensively in Utah's communities and intensively in Utahns' lives, then social divides and cleavages are more likely to be addressed with significantly less direct violence and its associated costs than otherwise would be the case. On the other hand, students will also witness what happens when public and civil society institutions have diminishing presence in the communities of Utah. That is, they are much less effectively addressing political issues of gender, race, ethnicity, class, and ideology. As a result the social divisions, cleavages, and conflicts in our communities will result in a significantly increased likelihood for direct violence, and direct violence will materialize on a much more systematic basis.³

On the other hand, the existence of vibrant public and civic sector organizations operating regularly in communities and citizens' lives requires particular things. These include wealth and the deliberate allocation of significant amounts of social resources to the extant infrastructural and institutional support networks across and within Utah's communities. But without the deliberate allocation of wealth and resources from both the public and private sectors, the public and civil society institutions, nongovernmental organizations (NGOs), organizations, and associations daily ministering to those in need as a result of

³ In the sub-field of Political Science known as Comparative Politics, these social factors and the divides they engender are called cleavages.

the social divides, cleavages, and conflicts in our communities will have a diminished ability to address such conflicts. With such decline, our communities become increasingly impoverished. With increasing social poverty comes increasing amounts of direct violence. Each feeds the other in a feedback loop of social disintegration, and the loop is intensified when there is economic collapse. That is, the number of social conflicts increase while the ability of the public and civil society organizations, institutions, NGOs, etc., to cope with them is steadily eroded through massive private and public sector budget cuts.

An empirical indication of this was the 'Great Recession' that formally began one month into my research [9/15/08, with the Wall St. collapse of Lehmann Bros.]. Virtually all interview subjects said that as a result of the economic collapse, their public and civic organizations had significantly diminished dollars (USD) and therefore abilities to directly address increasing numbers of social conflicts in Utah's communities.

This was especially acute amongst the DV shelter directors and associated DV advocates in Utah. All said that the need for DV shelter stays was not only increasing in number and frequency, but the average time DV victims needed to stay at any of the 16 DV shelters in Utah (funded by public and private sources) had extended beyond three months. These facts became evident in the months immediately following the collapse of the housing bubble and the subsequent Wall St. investment banking collapse [9/15/08]. It was, in other words, already quite ugly and getting more violent in and across Utah's communities, as a direct result of economic collapse.⁴

⁴ Two interview subjects told me in October, 2008, one month after the beginning of the 'Great Recession,' that as funds from the state government for their DV work were disappearing rapidly, their ability to run lock-down shelters for the victims of DV (mostly women and small children) and find them adequate shelter, on-going protection, and jobs in the larger community to support themselves and their children, was eroding steadily. This was only one empirical example discussed of the institutions of civil society losing their ability to minister daily and effectively to the victims of direct violence in the SLV. One immediate consequence of this loss was more homeless families living in and around Pioneer Park in downtown SLC. In other words, this community was becoming increasingly impoverished, and the directors of the civil society organizations spoke of it repeatedly. Another empirical indicator: the Utah Domestic Violence Coalition released a list of 22 Utahns who had been killed in 2008 as a result of DV; 15 of those resulted from the use of firearms. Taken together, these certainly are empirical indicators of disintegrating social conditions and circumstances, for more and more of the people of the SLV. As budgets in both private and public sectors have now [in 11/11] been sliced much

Being Flexible: ‘Learning By Doing’

Trying to gather ideas and information about the operations and activities of this large set of public and civil society organizations required flexibility. Experientially, learning while conducting the interviews meant questions had to be added as the interviews proceeded. As a result of additional queries pertaining to peace and justice work and how to best teach peace and justice to lower-division, higher education students, a subject suggested asking, “What is (peace and justice) activism?” Initially, this seemed valuable as interview subjects could help me compile a list of activities that would qualify as appropriate, non-violent forms of activism in support of peace and justice. But posing such a question could not only produce answers leading to the creation of a list, but also help establish some baseline in the determination of the costs that peace and justice work might regularly entail.

Stated differently, the U.S. has a Nobel Laureate (Joseph Stiglitz), some of whose recent written work seeks to account for the costs of directly violent conflict in the world (see *The Three Trillion \$\$\$ War*). Clearly, such authorship is intended to demonstrate the tremendous costs of the U.S. making war overseas, its national blood and treasure draining away in foreign lands.

On the other hand, given what my modest purpose was—researching and writing a PJS for SLCC—the inverse idea had utility. Creating a list specifying the costs of activities that rely on common strategies and tactics demonstrated to promote peace and justice in our communities and beyond might allow us to establish parameters concerning what each—i.e., waging peace versus war—costs in the 21st century. As a political scientist at SLCC, the idea of being able to make such a comparative cost accounting will undoubtedly prove extremely useful in the classroom with college students.

Another insight drawn from the research arose from the simplicity of the snowball method. Each interview was designed to allow a direct understanding of how and why different sorts of peace and justice work is done. Each subject was asked about the difficulties and details of their work. In addition, many of the interview questions posed were abstract in nature. These included, “What is peace, justice, and violence?” and “What would constitute justice in your field of work?”

Surprisingly, a significant number of interview subjects were not able to speak spontaneously and effectively about that which would

further than in 2008-09, Occupy Wall Street encampments have dotted the landscape (and the global one too), including Salt Lake City.

constitute justice in their fields. This is noteworthy at least because it reflects how complicated the fields of violent conflict in this community and well beyond are. But this ineffective articulation—most commonly from the directors of public and civil society organizations and institutions—in communicating the direct concerns of justice in their own professional fields also reflected their disjointed and disparate work efforts in the world. The work that is being done by numerous public and non-profit organizations functioning in civil society to make this a less violent world is being done mostly in separate silos. Such a disjuncture is an enormous cost for the greater SLV and well beyond, and like the winter inversions here, remains officially unaccounted for.⁵

⁵ To put this matter in political and economic terms, what is taking place in both circumstances is an externalizing of the costs of production. In the first, because of the decreasing abilities of humanitarian and NGO work (publicly and privately) to be well-coordinated in some central fashion—at the local and global levels of analysis and work, each operates in a separate manner while providing its civil society function. But there are few strong, central links between and among the organizations to promote their common interests and operations. As a result, violence in communities predominates more than do the organizations' activities and operations seeking to stop violence in our communities. As for the second, because of the 'automobile way of life' in the SLV, when the winter high pressure systems build over Utah, the air at the bottom of the valley gets filled and filthy with the exhaust of millions of internal combustion engines operating 24 hours 7 days a week. Asthma rates skyrocket among our children and others, and little is done to otherwise clean up the air we all must breathe. (Kudos to SLC mayor Ralph Becker for a public policy mandating otherwise.) The one field of peace and justice work where there seemed to be more coordination than all the others was the DV work being done by the women of the SLV.

One final point about costs being externalized, this time to say that most peace and justice work remains unaccounted for. That is because far and away the work of most of the organizations was done by volunteers. There is no reliable way to account for what is the value of all those volunteer hours. The non-profit sector nationally as of late 2008 was said [Reuters] to be worth 8% of the U.S. economy; however, even that statistic cannot account for any of those volunteer hours. Unfortunately, in this society when things are not accounted for in formal, quantitative terms, then they usually function at the margins of society. This can certainly be said about all the local forms of peace and justice

Teaching Peace and Justice Studies Requires Working 'Against the Grain'

In spring 2008, I wrote the sabbatical proposal that SLCC's administration accepted. The most significant personal and professional matter, therefore, was not only writing the courses after conducting the interviews, but actually putting a PJS into SLCC's formal curricular offerings. With that accomplished would follow the intentional teaching and learning of active and direct non-violent methods for addressing conflict, and the foundational reasons why this is in everyone's political interests.

Critically, with an SLCC PJS program in place, students and faculty would be working 'against the grain' (for a fuller accounting of what is at stake when we operate 'against the grain,' see Eagleton 1986). This would necessitate that a PJS curriculum and its participants regularly examine alternative, far less costly forms of thought, rationality, and activities than are customarily practiced in and by the U.S. That is, they must be willing to work against the logic of a society predicated upon a petrochemical infrastructure that is always preparing for war, presently making war, and still recovering from its last wars. In so doing, SLCC's PJS program would tap into the deeply felt need many students have expressed to me over the last 10 years for the political education and knowledge necessary for direct, non-violent civic involvement that can stop this madness. Such an education and these practices operate, furthermore, well beyond the thin democratic practices presently and usually available to American citizens.

The written SLCC PJS program was designed to incorporate service learning, so that non-profit, civil society and public organizations in the SLV could focus student attention on identified social, political, and economic needs in the larger community. Service learning, in other words, was the curricular method for leading the community college's greatest asset—its student population—into the organizations doing peace and justice (studies) work in the greater SLV. Thus, if a PJS program were in place, perhaps it could offer students an Associate of Applied Science degree in Community Organizing, as well as a Certificate in PJS. In turn, such accomplishments for SLCC's students might translate into paid work at the many NGOs and non-profit agencies, organi-

work this author encountered during the sabbatical year. In other words, citizenship in the U.S. rarely pays us to do good things. The SLCC PJS program has other ideas.

zations, and associations that now (and in the future) address the identified needs of the SLV.

Also working ‘against the grain’ is the written content of the two Peace and Conflict courses at the base of the SLCC PJS program. Each is designed to focus student attention on the examination of political conflicts that manifest regularly in direct violence, thereby increasing the costs for all actors involved. By definition, this also means that collateral costs inevitably fall on those in proximity whenever direct violence erupts, such as the consequences that gang violence and drug wars have in communities across the U.S., Mexico, and/or throughout the world. This is also the case with DV here and in the wider world. It is not surprising, therefore, that direct violence at the individual, domestic, international, and global levels of analysis share similar qualities and costs, especially in terms of who and/or what are the regularized winners versus the systematic losers of violent conflict.⁶

⁶ Here are two other glaring examples of the associated costs we all face when U.S. political leaders make war on foreign peoples, and which demonstrate the critical necessity for teaching ‘against the grain’: a) U.S. President George W. Bush’s suggestion to the American people soon after 9/11/01 that they should go shopping is illustrative of U.S. political leadership telling the American people to do what they already know how to do best. One unspoken message—at the same time—was that the most significant operations of the U.S. nation-state—especially in times of crisis—will be determined by those at the head of the U.S. government. Full democratic governance must be moderated or dispensed with altogether in times of crisis. In other words, the legislative arm of government (*viz.*, the Congress), the first institution of the U.S. Constitution, that which writes the laws and directly represents the interests of 310 million people, has little to no function anymore. Only a ‘unitary executive’ can act decisively—with dispatch, Alexander Hamilton said—to protect the nation’s security. Thus, on 9/11/01, there was no plan or design to hide the leadership of Congress, but complete plans and actions carried out to hide Dick Cheney and the other high priests of the executive [*viz.*, Bush II] administration; and b) It is worth noting another version of the consequences of the attacks of 9/11/01 can be deduced that operates ‘against the grain’: *viz.*, that the president and the U.S. presidency, which wields the greatest military machinery the world has ever known, was unprepared and unable—when it mattered most—to use that military machinery to protect the nation’s security. As a result, being a commander in chief who saw nothing coming on 9/11/01, President Bush actually failed to do what most consider his most significant task as the chief executive and commanding officer of the U.S. *i.e.*, protect the citizenry from such massive attack as happened on 9/11/01. Tragically, that is not, nor has been, among the messages directed at/to the American people from their government or the domestic, mass, commercial, corporation-owned media ever since the events of 9/11/01.

Costs of Organizing for Peace and Justice

Part of the process of writing a PJS curriculum included doing peace and justice activism in the SLV. This meant helping to organize the Second [Salt Lake] Winter Conference with Utah Jobs for Justice; the Wasatch Coalition for Peace and Justice [WCPJ]; and the Healthy Planet Mobilization Committee ["Strategy and Tactics of Organizing a Progressive Social Movement," 2/13-14/09]. Still more fieldwork included trips to collaborate with Stan Taylor, Ph.D., director of the Peace Center at Lane Community College [LCC, Eugene, OR]. This section summarizes the costs of the work of both entities, as each organized and held peace and justice conferences over two years [2008 and 2009]. The point was for me to learn how to organize such an event, as it was likely that a formal SLCC PJS program would itself organize and execute a peace and justice conference annually.

Both LCC's two Peace Conferences [3/08 and 5/09] and Salt Lake's two Winter Conferences [2/08 and 2/09] took place over two days. The *n* is low, but one illuminating difference was the gap in costs between the academic-led conferences done by the Peace Center at LCC and SLC's community-led activist organizations' two conferences. That is, LCC's conferences cost more than twice what the activist-led conferences here in Salt Lake City cost to organize and put on (\$8,700+ vs. \$3,000).

The large differences in cost between holding the academic-led peace and justice conferences and activist-led ones was reflected as well in the different internal dynamics of each. For example, the [SLC] Winter Conference in 2009 focused on promoting three political issues and brought invited speakers and presented multiple films to speak directly to those issues. The conference organized around and spoke for: 1) a single-payer health care system in the U.S.; 2) stopping the U.S. "Global War on Terror," especially the U.S. wars in/on Afghanistan and Iraq; and 3) U.S. policy-makers using national political power to address global climate change and design alternative energy systems in the U.S. The activist-led conferences had the freedom, in other words, to practice political advocacy.

The academic-led LCC Peace Conference in May 2009 was driven by a volunteer steering committee representing all of LCC's different employees. The committee chose to invite people to speak and perform at the conference who have been doing various forms of peace and justice work across the U.S. The speakers brought in weren't fit to a specific peace and justice agenda, but instead the conference agenda coalesced around the invited speakers and the activism and issues on which they work. From the keynote and plenary speakers to the spoken

word and musical performers, each directed audiences to various peace and justice realms, including racial justice; environmental justice; GLBT justice; economic justice; anti-war and pro-peace activism; Native American spirituality and peace; white privilege workshops; Iraqi refugee matters; and sustainable agricultural practices.

The academic conference thus acted as a networking forum for different community leaders around the country doing various types of peace and justice activism in their own organizing work, their art and poetry, climate activism, etc. In other words, LCC's two conferences—as a deliberate means of bringing together and engaging audiences around peace and justice work—relied on invited speakers, music and spoken word performers, and break-out sessions led by other invited leaders to establish the conference focus. In May 2009, approximately 800 people were drawn to attend the LCC Peace Conference.

On the other hand, the two activist-led conferences in SLC operated on shoestring budgets, including coordination of the means and methods of paying for essential items. They relied on volunteer labor for everything to make the conferences materialize: meeting every week for 7+ months to plan the conferences; creating the conference websites and the conference poster materials; putting up flyers; making public service announcements on/in local public radio and print; and ensuring that all the practicable matters functioned effectively as means of communication to the larger SLV community in which the two conferences unfolded. This also included local activists making use of other activist friends elsewhere, who came to town to speak because airfare and lodging were covered, and they needed no or little honoraria; however, the most significant item for which in both years there was no cost for the conferences here was renting facilities. This is because a few of the SLC activists from the three community organizations were employees of the University of Utah and made use of the University's facilities at no charge. As a result, both local activist-led conferences organized and held at the University of Utah cost around \$3,000. Altogether, it was estimated that at both activist-led Winter Conferences about 200 people attended.

Academic-led conferences, however, usually operate as professionally as the faculty organizing them can afford. Because academics are trained to rationalize things, and conferences are one such thing rationally organized on a regular basis in higher education, it could be argued that the mere professionalizing of the endeavor results in higher costs being administered.

Activists, on the other hand, often have to work with what limited financial means are obtained locally (no grants were written or funded for either SLC Winter Conference), but this also allowed a freer ex-

pression of commitment to specific political goals that the academic-led conferences did not offer. The activists operated cheaply, but in a much more traditional way, in having organized two peace and justice conferences. The academics operated 2+ times as expensively, but much less traditionally, for two peace and justice conferences. The latter also produced four times the audiences with that much greater audience participation, than the traditionally organized, activist-led peace and justice conferences.

All of these costs, conference organizing strategies and tactics, etc., are of the utmost significance, if there ever is to be an SLCC PJS Program. It would, by definition, need to hold a peace and justice conference at SLCC annually to reinforce the PJS curricular work and its desired student learning outcomes, especially civic engagement. Among many components of what it means to be engaged in civil society is a required amount of time individuals practice being connected to a whole greater than themselves. Such a process—by definition—calls for some regularized amount of individual sacrifice, which the next section highlights in converse fashion.

‘The Will to Power’

The will to power can be described as the belief that “I am right” and others need to follow “my” opinion about (political) things. This is something each (modern) human has inside him- or herself, although the best of the peace and justice workers interviewed showed little to no sign of it. On the other hand, certain subjects in the study did reflect the will to power to an extreme extent. That is especially pertinent here because others who have suffered from violence (direct or otherwise) will not be well served, e.g., by an NGO director who puts his or her own interests first.

Early in my discussions about peace and justice, I spoke with a member of Utah’s executive office. We conversed for almost two hours, and after a pause, this person said to me that the promotion of peace and justice would work out “as long as the abortionists and homosexuals didn’t interfere.” I was taken aback by this proclamation and responded: “I don’t know, but if people were keeping me in the closet when discussions about peace and justice in this community and beyond were being held, I’d be making a lot of noise.” Somehow the discussion ended and I left the office, but the deep impression left from those words will stay with me a long time. Put simply, that was the expression of bigotry by a person in a very powerful office, and it speaks of the will to power in institutional proportions, not just in individuals.

As already noted, each of us has some amount of the will to power. In this specific case, the will to power was clearly seen in ideas and speech that sought to define and exclude who (or what) can participate in the discussions necessary for the maintenance of political order, security, education, prosperity, etc. In turn, when someone in a public position of a very powerful office conveys that s/he knows who ought to be sitting at the peace and justice tables, that person is operating from an ideological position not supported by reason. However, that person has the power to help ensure which individuals and groups will have seats at the table and which won't, and that is deeply problematical in democratic politics. The will to power of this individual can become political outcome, contrary to (all) people ruling themselves. That such a will to power sat at the top of the executive office of the governor of the state is not acceptable in/to any PJS curriculum.

The second incident of a subject expressing an extreme will to power was an interview of one of two war veterans in the subject population. We had met a few weeks earlier at an organizing meeting for the WCPJ. I said I'd like to interview him for my sabbatical project and over the phone thereafter, we agreed to meet over dinner.

Again, there is the belief in of each of us that "I know best" what needs and/or ought to be done [politically] in the world, but this interview subject talked so much that he never even asked about my questions. I couldn't get a word in edgewise; dinner was spent with him talking incessantly about his plans for putting the world in proper order. His plans were getting elected to the U.S. Senate from Utah in 2012 and then to the U.S. Presidency in 2020. His confidence was high, but this seemed most to reflect delusions of grandeur.

Here again, however, the will to power appeared in an individual who believed he knew how to properly order the world. In addition, I was told he planned on putting all his energies toward being in a position to make himself electable to do so. He was not interested in any of my questions. He was a U.S. military veteran and had been in the 'belly of the beast' (both in Iraq and the Pentagon's bureaucracy), has a Master's degree in accounting, and as a result believed he knew best what needed to be done [politically] to put the world to right.

The straightforward principle learned from both experiences was the certainty that a SLCC PJS Program would always have to be on guard against the will to power. With such examples, it is not hard to see that the will to power is more characteristic of a partial set of ideas [viz., ideology] seeking dominance, rather than the reason of an educator seeking critique. Conversely, a PJS has to be led by principles of inclusion, respect for difference, listening to others, and a commitment

to human rights that promotes a sustainable quality of life for all peoples everywhere.

As opposed to the will to power, the SLCC PJS curriculum was written as a reasoned approach to examining, describing, and explaining conflict, especially how and why direct violence is often enough involved, at the individual, domestic, international, and/or global level of analysis. Part of this reasoning includes the empirically grounded assertion that conflict at all levels—with and without direct violence—seeps from one realm regularly to another. To say otherwise, e.g., about the popular culture of the U.S. at a global level, is simply foolish.

In addition, with its four current wars abroad (Iraq, Afghanistan, Somalia, and now Libya), U.S. foreign policies are involved in directly violent conflicts in the world like no other single international actor (state or non-state). Indeed, the U.S. National Intelligence Estimate for each year following President G.W. Bush's war on Iraq [3/19/03] has shown that al-Queda-linked and/or -inspired terrorist attacks (especially on western [viz., U.S.] targets) have increased worldwide, even as the U.S. has had its soldiers in over 1,000 locations globally, across 130+ nation-states, with 730 official bases beyond the borders of the U.S. Having this global military presence, however, only seems to increase the desires of others to question U.S. legitimacy and its operations on their soil. They seem to be acting regularly with direct violence against the stated and practiced interests of the US abroad.

Thus, at the heart of the will to power is the belief—and not infrequently the power—that an individual and/or group has to dictate who/what has social standing to determine the significant social, economic, and political matters of the domestic and/or international realms. In the PJS Certificate Program, however, this is a large and complex question open for student deliberation.

Victimization First, Empowerment Afterwards: 'A Political Economy of Direct Violence'

Another lesson learned during the interviews came from the Diversity Services Coordinator for the SLC YWCA. The Diversity Services Program began in 1994. It has provided a privileged space for people of color who are victims of DV. Victims of DV get translators, crime advocates, legal representation, etc. Like the Family Justice Center (the YWCA's building next door), it has been a lock-down facility for victims of DV. The focus of this organization has been on female victims of DV who are refugees and immigrants, but they do service male DV victims too (although this is rare). The coordinator saw her work efforts as empowering women, by breaking language barriers and

acting as cultural advocates for her clients. Her business card said, “empowering women, eliminating racism” (which is the YWCA’s slogan). I asked her how she believed her job (and the YWCA itself) was accomplishing what the business card claimed. She responded that they did this “one victim at a time.”

Her insight in this single statement was powerful. It said descriptively that each woman first has to be a victim of abuse, violence, etc., before she can become empowered to actively resist racism. As a result of being a victim of direct violence, in other words, each has a chance of learning [from the YWCA’s gifted employees] how to wield more power in a world that might include less racism. Again, this could only take place after the abuse and violence against her clients had happened. At stake in this upside-down, counter-intuitive depiction of DV and its aftermath seemed to rest complementary but also oppositional variables and values. That is, must formerly powerless people first be victimized by direct violence in order to become politically empowered in the world?

To answer this question, in this last section, I now turn the will to power inside-out, dissecting it to construct ‘a political economy of direct violence.’ It starts with the following commonplace notion: Conflict is inherent in the human condition. However, conflict per se and directly violent conflict are distinctly different phenomena, which ‘a political economy of direct violence’ seeks to describe and explain. It follows:

- 1) Whenever direct violence is first introduced into a conflict the costs of the conflict rise immediately and precipitously.
- 2) These costs are not only projected onto the immediate and primary persons involved in the conflict that has turned violent, but as the Pentagon likes to say about all of their wars in the late 20th and early 21st centuries, these costs [of direct violence being introduced into a previously non-violent conflict] will be externalized onto innumerable collateral victims.
- 3) Consequently, whether direct violence is first introduced into a conflict by a criminal gang, the state, or an abusive parent or spouse, the outcome of the introduction of direct violence into a conflict is negative sum for all involved. Thus, in one way or another, everybody has to bear costs not previously at stake in the conflict.
- 4) Although everyone who (and everything which) faces the vastly increased costs of the newly arranged conflict loses, these costs are nonetheless distributed (and are borne/carried/faced) in ways that usually cost the violent actors the least, particularly in the short term.

- 5) In turn, to use direct violence first is to seek—by definition—to dominate others, and this is so simply because most other people[s] (and most of their social organizations and institutions) cannot afford the immediate and associated costs of the direct violence. But as this states, even the violent must accept costs they previously had not. (Becoming violent always has immediate costs for the violent, to say little of the inevitable repercussions, whatever they may be.)

As ‘a political economy of direct violence,’ this is proposed as a basis for a model as to how domination and subordination is established and sustained in (post-)modern times. Moreover, it is worth complementing with other ways of effectively describing and explaining how and why the directly violent operate as they do. The crux therefore is that any PJS curriculum must examine and explicate the costs and consequences of violence, if college students are ever to be prepared to actively work toward stopping it in the world.

Here, then, is a second description of a political economy of direct violence. First, we must differentiate between conflict and violent conflict, because the rationality involved in each type of conflict is different: a) to use direct violence in any conflict previously nonviolent is to literally “up the ante” or “raise the stakes” of the conflict; b) as a consequence of the introduction of direct violence, the party/parties to any conflict who were/are the targets of another’s party’s violence, now have entirely different decisions to make about themselves. Most significantly, what will they do in response to the direct violence just imposed on them?

If they choose to respond in kind with direct violence targeting the initially violent actor, that action further raises the costs of the conflict, and does so in exponential fashion. This is because a response of violence to the initially violent requires one of two things: Either it is so overwhelming in force to the initiator that the initiator ceases being violent, or the violent responder is unable to do so, allowing the initially violent to then respond with more and greater violence. In other words, each directly violent move must seek to be preponderant and/or overwhelming over the opponent[s], or it hasn’t the likelihood of effectively stopping the violence in any actors’ interests.

Second, all victims to newly introduced violence have greater costs to bear right now, in the immediate term. That changes everything for all parties involved in a newly violent conflict, including the party [or parties] that introduced the violence. As a result, conflict and violent conflict do not involve the same dynamic, and that is why all is fair in war. In war, the rule of law does not apply. As a result, in war, for-

eign and/or domestic military forces regularly kill civilians, and there are no courts, magistrates, or police to stop this from happening.

Third, it is significant to note that the first half of the aphorism—that all is fair in love and war—is simply false in the (post-)modern world. If I start to beat my spouse, or vice versa, the community should call that DV, and that's a crime. By definition, it's illegal for one person to beat another; when it isn't DV, then it's often called assault and battery. The laws and the police, in other words, recognize the difference between conflict *per se* and violent conflict. Also, our community's social norms recognize the distinction between conflict and directly violent conflict. Again, the larger point is that there is conflict *per se* and then there is directly violent conflict. The two are distinctly different.

Fourth, the U.S. Constitution allows us to disagree, and to do so forcefully in the public sphere to get our various messages across to the people of the nation. This has recently been made clear by the U.S. Supreme Court, all based on the First Amendment's right to free speech [8-1]. Any people who are organized, acting in political and peaceful ways, have the right to broadcast their speech and have the right to do so where their message must be heard. This was as true of the variety of protesters here in SLC for the Winter Olympics in 2002 and for the Westboro Baptist Church (WBC) from Kansas. They used these rights in SLC at West High School back in early February 2011. They were vociferously picketing that public high school's gay-straight alliance organization of students. The WBC protesters held signs saying, among other things, "God Hates Fags."

However, what is being said here about conflict *per se* is that disagreements, conflicting points of view, independent entities like nation-states, separate religions and their followers, all of these sorts of discrete things expressing varied and clashing points of view in the world are to be expected. Conflict is inherent in the modern human condition, but unfortunately and often enough, conflict becomes directly violent conflict. On the other hand, the difference between the two types of conflict is empirically verifiable and is recognized by the U.S. Supreme Court as essential to our liberties, and this is because conflict *per se* and violent conflict operate with different dynamics involved and different costs at stake.

Finally, to put this matter conversely, violence, by definition, is not conflict. A synonym check in Microsoft Word for the two concepts reveals that although a few violent words are included as synonyms for conflict, but these few are among many others that do not denote violence (e.g., disagreement, divergence, dispute, quarrel, argument). However, VIOLENCE has no synonyms (among many) that include

the word CONFLICT. Thus, the two things—conflict per se and violent conflict—are distinct denotatively.

Conclusion

There is no definitive reason why formerly powerless people must first be victimized by direct violence to become politically empowered in the world, actively working against direct violence [e.g., racism]. As these two inter-related lines of analysis tracing ‘a political economy of direct violence’ demonstrate, to use direct violence first in any conflict is to deliberately seek to subordinate another (and/or others). This can be seen simply by comparing what the costs are and how they are borne, when using direct violence versus using active, non-violent means of addressing conflict. This is so in the immediate and any long-term, and at all levels of political analysis.

As already stated, the introduction of direct violence into previously non-violent conflicts has quite similar dynamics at the micro- and macro-political levels of analysis. That is, the production and distribution of the costs of direct violence follow similar patterns and processes, whether at the level of DV within a family (the micro-cosmic, individual level) or at the international and/or global level of analysis (e.g., when the U.S. imposes “shock and awe” on Iraq on 3/19/2003). In addition, instances of direct violence occurring at all levels of political analysis share identifiable and common characteristics worthy of attention. To date, very little analysis and attention on these empirically similar characteristics has been done in political science in the U.S. That too costs us all. However, with the establishment of an on-going SLCC PJS Program, some of those costs might someday be addressed and accounted for, especially with the introduction of ‘a political economy of direct violence’ in the curriculum.

References

Bilmes, Linda, and Stieglitz, Joseph. *The Three Trillion Dollar War: The True Cost of the Iraq Conflict*. Reed Business Information, 2008.

Eagleton, Terry. *Against the Grain: Selected Essays*, Verso, 1986.

Engelhardt, Tom. “The US Has 761 Military Bases Across the Planet, and We Simply Never Talk About It,” <http://www.tomdispatch.com>, September 8, 2008. Accessed on April 13, 2009, at <http://www.alternet.org/story/97913/>.

National Intelligence Council. http://www.dni.gov/nic/NIC_home.html.

SNYDER v. PHELPS ET AL., [US] Supreme Court Case No. 09–751. Argued October 6, 2010, decided March 2, 2011.

Von Ahn, Lisa. “Investment adviser asks if wealthy are necessary,” <http://www.reuters.com/article/businessNews/idUSN1140057120070915>, September 15, 2007. Accessed February 21, 2012.

The Regressives: Unraveling the Progressive Movement

John Howell

Southern Utah University

Abstract

The Progressive Movement of the late 1800s and early 1900s was one of the most influential political movements in American history. Progressives sought to shift power away from political elites and into the hands of the citizenry. Progressives viewed more democratic government as both less corrupt and more responsive to the will and needs of the people. The early Progressives were very much in line with the Populist movement of the same time period. The modern political landscape has many who claim the Progressive label, yet have abandoned their belief in many of the tenets of the Progressive Movement, while others have been retained. The goals of modern Progressives are more often pursued through the very elite powers that the original Progressives opposed, often in direct opposition to the stated wishes of "We the People." In effect, they have become that which their ideological ancestors despised.

American political history can roughly be divided into a series of eras, each dominated by its own particular groups and personalities.

Some notable American eras include the Era of Good Feeling, the Jacksonian Era, and the Gilded Age. Each era can be seen to some extent as a reaction to that which came before, often drawing upon the laudable idea of curbing the excesses or shortcomings of the preceding era. Such is the case with the Progressive Era of the late 1800s and early 1900s. The Progressive Era can be seen as a reaction to the Gilded Age, which was an era of nearly unbridled free-market capitalism and the accumulation of almost unimaginable wealth. Men such as Andrew Carnegie, Andrew Mellon, and John D. Rockefeller made their fortunes during this time. Along with this great concentration of wealth came a desire by powerful people to influence government, usually through means that were fantastically corrupt by today's standards. The Progressive Movement sought to curb the influence of political elites and those that sought to control them by shifting governmental power away from political elites and the wealthy and into the hands of the citizenry. Notable successes of the early Progressive Movement in this endeavor include direct popular election of United States Senators with the passage of the 17th Amendment, the expansion of suffrage to include women with the 19th Amendment, efforts to limit the influence of money in politics, the direct primary as a means of selecting presidential nominees, the ballot initiative that allowed disgruntled citizens to bypass intransigent state legislatures, and the introduction of the secret ballot to inhibit political intimidation of voters (Baradat, 2003). The last three efforts are the subject of this paper. Each will be presented as a case study, along with modern efforts by those claiming the title of Progressive to reverse the perceived beneficial democratic effects of the Progressive Movement.

It is important to note that there is no direct link between the Democratic Party and either the original or modern Progressive Movement. When the movement began, there were prominent members of both parties who were considered Progressives. Theodore Roosevelt and Robert LaFollette can rightfully be called Progressive Republicans, and notables such as William Jennings Bryan can be cited as Progressives from the Democratic Party. Today, while the Democratic Party may be more in line with Progressives, it is certainly criticized by them. Many modern Progressives vilify the Democratic Party as being as much of an impediment to their goals as the Republican Party. In short, Progressivism should not be considered as synonymous with the Democratic Party.

The 2008 Democratic Party Nomination

The 2008 Democratic Party nomination battle was historic in a number of ways. Most obviously, it pitted a female candidate against a

black candidate. Neither a female candidate nor a black candidate had ever come close to winning a major party's presidential nomination, and once the race was effectively down to two candidates, it became obvious that, one way or another, that was going to change. Less obviously but also importantly, it was the first nomination fight in which the Democratic Party's "superdelegates" became a possibly deciding factor (MSNBC.com).

1968 and its aftermath

1968 is rightfully considered a defining year in American political history. America saw a president who was eligible for re-election choose not to seek the nomination, which is exceedingly rare. President Lyndon B. Johnson's stunning announcement on March 31, 1968, that he would not seek the Democratic Party nomination threw the battle for the Democratic Party nomination wide open. Leading candidates Senator Robert F. Kennedy of New York and Senator Eugene McCarthy of Minnesota were both in strong opposition to Johnson's Vietnam War policy and pledged to bring a quick end to American involvement in Southeast Asia. When front-runner Robert F. Kennedy was assassinated on June 5, 1968, it seemed apparent that convention delegate leader Eugene McCarthy would win the Democratic Party nomination. Vice president Hubert Humphrey had officially entered the race on April 27, but most assumed that McCarthy would win the nomination because of his huge lead in convention delegates. Humphrey was also burdened with his close association with President Johnson, whose favorability ratings were in the mid-30% range. A majority of Kennedy's delegates shifted their allegiance to Humphrey, however, so that when the Democratic National Convention roll call vote was taken, Humphrey won easily. (Witcover, 2003).

The 1968 Democratic National Convention in Chicago was, to put it mildly, chaotic. Unrest was evident both on the convention floor and on the streets. Outside the auditorium, at least 5,000 protesters clashed with police, and television cameras were there to record the aggressive response of Mayor Richard J. Daley's police. Inside the auditorium, delegates chose to follow the directives of state party organizations rather than nominate McCarthy, who had done well in the primaries. Choosing the will of party bosses over the voice of Democratic primary voters, Humphrey won the nomination and succeeded in retaining Johnson's controversial Vietnam War policy in the party platform despite strong opposition (DeGregorio, 2005). Humphrey's eventual narrow loss to Republican Richard Nixon left deep wounds in the Democratic Party, mostly because of the ugly nomination skirmishes.

Humphrey's nomination over Eugene McCarthy, who had won more delegates in the Democratic Party primaries, prompted calls for reform in the nomination process. State party primaries to select national convention delegates were not a new idea, having been one of the Progressive reforms pushed by Wisconsin's Robert LaFollette in the early 1900s (Jackson and Crotty, 2001). The Progressives believed that direct primaries would combat the cronyism and "smoke-filled room" corruption that had become all too common in the nomination process. "In the view of Progressives, the primary would strike at the very heart of the power exercised by political party leaders, that is, their control over the nomination process. If party leaders can control who the nominees will be, they have gone a long way toward determining what the final result will be." (Jackson and Crotty, 2001). As with their other proposed reforms, Progressives sought to shift power away from political elites and toward the people. Wisconsin, as a state that was prominent in the Progressive Movement, held the first direct primary to nominate party convention delegates in 1904 (Lovejoy, 1911). Other states followed in both the Democratic and Republican Parties:

Table 1 indicates a fairly steady number of states holding primaries to select convention delegates as well as a steady percentage of delegates chosen by primary from 1912 to 1968. Both numbers increased dramatically in the aftermath of the 1968 Democratic National Convention. Reforms were instituted, mainly through the Democratic Party's McGovern–Fraser Commission, which fundamentally changed the nature of the relationship between the national party organizations and the state party organizations. The McGovern–Fraser Commission's most important tenet was that all registered Democratic voters in the states should have 'the maximum feasible opportunity to participate in the delegate selection process.'" (Jackson and Crotty, 2001). In response to the rules promulgated by the McGovern–Fraser Commission, many states adopted the direct primary as their method of selecting national convention delegates. Republicans noted the changes the Democrats were making and followed their reforms with similar initiatives outlined by the Delegates and Organization Committee (Jackson and Crotty, 2001). It seemed that after nearly three-quarters of a century, the Progressives had their desired reform.

But did the Progressive reform of shifting power away from party elites and back to primary voters suffer a setback at the 2008 Democratic National Convention? An analysis of the delegate count supports that contention. Although estimates varied according to the source, a candidate needed 2,025 delegates to obtain a majority and win the nomination. As the convention approached, neither Barack Obama nor Hil-

Table 1. Presidential Primaries, 1912–2008				
	Democratic Party		Republican Party	
Year	No. of primaries	Delegates selected in primaries (%)	No. of primaries	Delegates selected in primaries (%)
1912	12	32.9	13	41.7
1916	20	53.5	20	58.9
1920	14	44.6	20	57.8
1924	14	35.5	17	45.3
1928	16	42.2	15	44.9
1932	16	40.0	14	37.7
1936	14	36.5	12	37.5
1940	13	35.8	13	38.8
1944	14	36.7	13	38.7
1948	14	36.3	12	36.0
1952	16	38.7	13	39.0
1956	19	42.7	19	44.8
1960	16	38.3	15	38.6
1964	16	45.7	16	45.6
1968	15	40.2	15	38.1
1972	21	65.3	20	56.8
1976	27	76.0	26	71.0
1980	34	71.8	34	76.0
1984	29	52.4	25	71.0
1988	36	66.6	36	76.9
1992	39	66.9	38	83.9
1996	35	65.3	42	84.6
2000	40	64.6	43	83.8
2004	38	83.2	27	56.9
2008	39	83.1	39	55.2

Source: *Vital Statistics on American Politics, 2007–2008*, (Washington, D.C.: CQ Press, 2008).

lary Clinton had the needed number of primary-selected delegates. Obama's winning margin was achieved not through the delegates chosen by primary voters but by superdelegates.

The Democratic Party instituted the super delegate system in 1984. Unlike with the changes following 1968, the Republican Party has not responded with a similar system. Superdelegates are elected members of the Democratic National Committee (DNC), Democratic governors, Democratic Senators and House Members, distinguished party leaders, and unpledged delegates chosen by the DNC. The rationale behind superdelegates was twofold. First, superdelegates are designed to act as a check on ideologically extreme or inexperienced candidates. Second, superdelegates give power to people who have a vested interest in party policies: elected officials. Put simply, superdelegates are party elites who are not chosen by voters in party primaries, and their interests may differ significantly from the interests of the primary voters (<http://www.definitions.uslegal.com/s/superdelegate>). In 2008, there were approximately 850 superdelegates at the Democratic National Convention. Does Democratic candidate Barack Obama owe his nomination to the more progressive primary delegates or to the party elite superdelegates? Obama certainly could not have won the Democratic Party nomination without the support of the superdelegates, who just as easily could have echoed 1968 and chosen Senator Hillary Clinton, even though she received fewer primary votes. While the eventual outcome reflected the will of the voters, the process looked suspiciously like the "bad old days" of smoke-filled rooms, party bosses, and "brokered conventions." The mechanism is certainly in place for just such an event.

The Employee Free Choice Act and the Secret Ballot

Among the Progressive reforms that were prominent in the early 1900s, the secret ballot is extremely important. Until that time, voting was generally public, and it was easy to determine how an individual voted. Along with this knowledge came obvious opportunities for intimidation from employers and others who had an interest in seeing that people voted "right."

Most political scientists trace the origins of the secret ballot to Victoria, Australia. While some forms of secret balloting probably occurred elsewhere, Australia was certainly the place where government first took on the responsibility of printing official ballot slips that were then handed out by election officials (Fredman, 2005). The secret ballot became law in Great Britain with an 1872 act of Parliament entitled, "An Act to Amend the Law relating to Procedure at Parliamentary and

Municipal Elections,” then made its way to America near the turn of the century (Bertrand et al., 2006). The importance of the secret ballot was touted by California reformer John Bidwell:

The Australian ballot system to me is the most important; I believe that no political party at this time appreciates its value...I almost believe that system is so important that the existence of our country depends upon it. I believe that if the people can express themselves, if they can exercise their intelligence, let it be intelligence small or great, if they can only exercise that intelligence – the intelligence itself will grow. (San Francisco Daily Evening Bulletin, March 25-26, 1891; San Francisco *Alta California*, January 15, 1891)

As well as having the effect of enforcing secrecy, which was a Progressive goal, this form of ballot also had the additional effect of expanding the electorate, which was another part of the Progressive platform. Bidwell’s view also reinforces the Progressive belief in a trust in the collective will and collective intelligence of the citizenry.

Disagreement does exist over whether the secret ballot, which was intended to promote democratic ideals and inhibit intimidation of voters, actually had the opposite effect. New York Governor David Bennett Hill, an upstate Democrat who worked closely with the New York City Tammany Hall machine, argued that the secret ballot inhibited the right of discussion and therefore the democratic process:

The system contemplated by this bill is a mongrel one, which has not been tested anywhere by experience... Here the people are upon an equality, and at the ballot-box all are freemen and equals. It has never heretofore been regarded as a crime for one citizen peaceably to discuss with his neighbour at the polls the merits of the various parties and candidates, and to compare views and to inform each other, if they desired, how they intended to vote... I realize that there is a class of well-meaning people who seem to hail with delight every new scheme which masquerades under the seductive name of ‘reform,’ especially if it comes from foreign shores and bears the approval of a monarchical government. (Fredman, 43)

In private, however, Governor Hill was clear concerning his partisanship. Hill knew that the current arrangement in New York encouraged straight party line voting, and that, along with the secret ballot

reformers were suggesting, ticket splitting had the potential to be harmful to the Democratic Party in New York.

The secret ballot has become perhaps the most enduring fixture of the Progressive Movement. Most Americans cannot conceive of a system in which they do not vote privately and in which their electoral wishes would be public information. Even in the election just past, stories abounded of intimidation of and violence against voters who expressed a preference through yard signs and bumper stickers. Neither side had a monopoly on such abhorrent behavior. It is difficult to argue that the problem would not get worse if an individual's vote choice was deemed to be public information. Other abuses, such as denial of employment or services, are certainly conceivable.

The Employee Free Choice Act (EFCA) is, at best, dubiously named. Under existing law, when workers are interested in unionizing, a private ballot election is usually held, and the election is overseen by the National Labor Relations Board (NLRB). The NLRB has established standard operating procedures to ensure free and fair elections, free of fraud and intimidation by either pro- or anti-union forces. The EFCA would effectively end a worker's right to a federally monitored ballot election with secret ballots. Instead, workers who favor a union would be asked to sign a card to that effect, in the presence of union organizers. Pro-union workers would be known to the employer, to co-workers, and to union organizers, as would those who decline to participate. The signing of this card is responsible for the term "card check" by which the bill is informally known.

The EFCA is a strongly partisan issue. Barack Obama supports it, telling the AFL-CIO, "We're ready to play offense for organized labor." (www.cnn.com/2008/POLITICS, 2010) John McCain opposes it, calling it "a poorly disguised attempt by the labor unions to swell their ranks at the expense of workers' rights and employers." (Reinhard, 2008). One attempt has already been made to pass the EFCA as HR 800. The House passed the bill on March 1, 2007, by a vote of 241 to 185. The vote was almost entirely on party lines, with only 13 Democrats voting against it and only 2 Republicans voting in favor of it (<http://clerk.house.gov/evs/2007/roll118.xml>). The Senate attempted to invoke cloture on the EFCA on June 28, 2007. The vote was 51–48, falling short of the necessary 60 votes. No Democrats opposed cloture, and only 1 Republican supported it (Bill Summary and Status, H.R. 800, <http://thomas.loc.gov/cgi-bin/bdquery/z?d110:h.r.00800>). With recent gains by Democrats in the 2008 election, prospects for passage of the EFCA seemed significantly improved. The EFCA has once again been introduced in the House and Senate and is currently at the subcommittee level.

Unsurprisingly, business leaders also oppose the EFCA. Bernard Marcus, co-founder of Home Depot, argues that the EFCA will make American companies uncompetitive in the global marketplace and ultimately send jobs overseas. Marcus also opposes the Act on political grounds: “The act, deceptively named the Employee Free Choice Act, eviscerates traditional democratic principles by effectively taking away an employee’s right to vote by secret ballot in union elections.” (*Bloomberg Businessweek*, 2008). Even liberal icon George McGovern opposes the EFCA. McGovern echoes the theme of this paper, writing that the measure “runs counter to ideals that were once at the core of the labor movement. Instead of providing a voice for the unheard, it risks silencing those who would speak.” McGovern also calls the bill “a disturbing and undemocratic overreach not in the interest of either management or labor.” (*Bloomberg Businessweek*, 2008).

The Citizen Initiative and California’s Proposition 8

Another pillar of the Progressive Movement was the citizen initiative or referendum. Through this process, which was and remains more common in the western states, citizens’ groups could effectively bypass a recalcitrant legislature and place prospective laws on the ballot for a direct vote of the people. Along with similar citizen-driven initiatives such as the recall vote, citizen initiatives reflected the Progressive Movement’s goal of placing more power in the hands of the voters and less in the hands of the elites.

In recent decades, California has been a leader in high-profile citizen initiatives. Memorable initiatives include Proposition 13, a 1978 initiative restricting property tax increases (California Taxpayers’ Association, 1993) and Proposition 209, a 1996 initiative approved by 54.7% of the voters of California that prohibited public institutions from considering race, gender, or ethnicity (Center for Individual Rights, 2006). Opponents of Proposition 209 challenged the initiative in court and were temporarily successful in blocking its implementation. In 2001, the University of California Board of Regents made a symbolic gesture in overturning their own ban on consideration of race, gender, and ethnicity in admissions decisions, but the state law remained in effect.

More recently, California has been at the center of the debate over same-sex marriage. In 2000, California voters approved Proposition 22 by 61.4% of the vote (Marriage Law Project, 2001). Proposition 22 changed the California Family Code to formally define marriage in California as between a man and a woman. On May 15, 2008, the California Supreme Court overturned Proposition 22 in a 4–3 vote, effec-

tively opening the door for same-sex marriages (“In re Marriage Cases,” 2008). The basis of the ruling was a violation of California’s Equal Protection Clause. While same-sex marriage proponents celebrated and marriage ceremonies were performed throughout the state, opponents succeeded in placing Proposition 8 on the ballot for the November 4, 2008, election. Proposition 8 was a constitutional amendment that sought to overturn the California Supreme Court’s ruling. In an extremely contentious vote, Proposition 8 was approved by a vote of 52–48% (Audi et al. 2008).

On November 5, 2008, the first lawsuits challenging the legitimacy of Proposition 8 were filed. A suit filed by the American Civil Liberties Union (ACLU) and Lambda, a prominent gay rights organization, contended that Proposition 8 actually revises the state constitution by altering such fundamental tenets as equal protection guarantees and that a measure to revise the state constitution can be placed before voters only by the Legislature (ACLU of Southern California, 2008). Proposition 8 co-chair Frank Schubert said, “This is exactly the type of behavior that brought to this position to begin with. The people voted eight years ago overwhelmingly in favor of traditional marriage, and they seem to be saying in pretty strong terms again ... that they favor traditional marriage, and yet this is not accepted by gay-rights activists.” (Richard Dawkins Foundation for Reason and Science, 2008).

On August 3, 2010, U.S. Chief District Judge Vaughn Walker ruled California’s citizen-approved constitutional ban on same-sex marriage unconstitutional, citing equal protection issues. “Proposition 8 fails to advance any rational basis for singling out gay men and lesbians for denial of a marriage license.” Walker wrote in his 136-page opinion. (*Perry v Schwarzenegger*, 630 F.3d.909 (9th Cir. 2011)).

The judicial system is recognized as the least democratic branch of government. Although more democratic than the federal court system because gubernatorial appointments must eventually be approved by a vote of the people, the California courts are certainly the least democratic branch of the state’s government. The federal courts are indeed the least democratic at all, with the citizenry having no electoral influence at all. In appealing to the courts to overrule a citizen initiative (Proposition 22) and to attempt to block another citizen-driven constitutional amendment (Proposition 8), opponents reveal themselves to also be opponents of power in the hands of the citizenry and proponents of elite domination.

Discussion and Summary

The most important conclusion of this paper concerns the status of those in the modern era calling themselves progressive. To put it clearly, these are not your father's (or your grandfather's or great-grandfather's) Progressives. The Progressives of the early 1900s were insurgents; they were the outsiders fighting the entrenched powers of corrupt government and big business. They trusted the popular will over the will of representatives, which they perceived to be interested in serving interests other than those of the general public. The Progressives of the early 1900s can be characterized as favoring "good government," which they saw as "government by the people" and potentially more responsive to the will of the people rather than the wishes of business and industry. In short, they were far closer to "small d" democrats than were the Framers of the Constitution.

Modern progressives share many of the same goals of their earlier counterparts. In the spirit of Progressive Republican Theodore Roosevelt, they still favor regulation of business. They continue to be advocates of the progressive income tax, usually advocating increased income taxes on the wealthy. They continue to share the goal of expanding the electorate, this time to include groups such as convicted felons and illegal immigrants (www.bamn.com/bamn-principles). The major difference between modern progressives and traditional Progressives lies in their belief in the will of the people. In each of the three presented case studies, it is the modern progressives that oppose more power being entrusted to the people. Modern progressives have become that which their namesakes of the early 1900s opposed: the entrenched power. While continuing to give lip service to "power to the people" and democratic ideals, modern progressives are anything but small "d" democrats. Modern progressives trust the will of the people exactly up to the point where the people no longer support their policy goals. When the will of the people does not support their aims, that will is quickly subdued through undemocratic means. This form of progressivism has shrunk from "power to the people," regressing to the same "power over the people" that has all too often been the hallmark of politics since the dawn of modern history.

References

ACLU of Southern California. Legal groups file lawsuit challenging proposition 8. November 5, 2008. <http://www.aclu-sc.org/releases/view/102912> (accessed 2/23/12)

Audi, T., Scheck, J. Lawton, C. "California votes for Prop 8." *Wall Street Journal*, November 5, 2008. <http://online.wsj.com/article/SB122586056759900673.html> (accessed 2/26/12)

BAMN Principles. <http://www.bamn.com/bamn-principles/>. (accessed 2/16/12).

Baradat, Leon P. (2003). *Political Ideologies: Their Origins and Impact*, 8th edition. Upper Saddle River, NJ: Pearson Education, Inc.

Bertrand, Romain, Jean-Louis Briquet, and Peter Pels. (2006) *The Hidden History of the Secret Ballot*. Bloomington: University of Indiana Press.

"Bill Summary and Status, H.R. 800." 2007. <http://thomas.loc.gov/cgi-bin/bdquery/z?d110:h.r.00800>:

California Taxpayers' Association. CalTax Research. Proposition 13: Love it or hate it, its roots go deep. November 1993. <http://www.caltax.org/research/prop13/prop13.htm>. (accessed 2/24/12)

Center for Individual Rights. Proposition 209 ends preferences in California. http://www.cir.usa.org/cases/prop209_info.html. (accessed 2/24/12)

DeGregorio, William A. (2005) *The Complete Book of the U.S. Presidents*, 6th edition. Fort Lee, NJ: Barricade Books, Inc.

"Employee Free Choice Act: Labor vs. Business." *Bloomberg Businessweek*. September 22, 2008. http://www.businessweek.com/bwdaily/dnflash/content/sep2008/db20080919_897469.htm (accessed 2/16/12).

Fredman, L.E. (1968). *The Australian Ballot: The Story of an American Reform*. East Lansing: Michigan State University Press.

Garrison, Jessica, Maura Dolan, and Nancy Vogel. Gay-rights advocates to challenge Proposition 8. <http://www.chicagotribune.com/news/nationworld/chi-081105-gay-marriage-ban-california,0,1804310.story> (accessed 11/6/2008)

"In re Marriage Cases," (2008) Supreme Court of California. Opinion No. S147999, filed May 15, 2008. <http://caselaw.findlaw.com/>

data2/californiastatecases/S147999.PDF

Jackson, J.S. and Crotty, W. (2001). *The Politics of Presidential Selection* (2nd. ed.). New York: Addison-Wesley Educational Publishers.

Lovejoy, A.F. (1911) LaFollette and the establishment of the direct primary in Wisconsin, 1890-1904. New Haven: Yale University Press

Marriage Law Project. 2001. California Proposition 22. <http://www.marriagewatch.org/media/prop22.htm>

“Organized labor divided on Clinton, Obama.” May 1, 2008. <http://www.cnn.com/2008/POLITICS/05/01/dems.unions/index.html> (accessed 11/4/2008).

Perry v Schwarzenegger. 2010. <http://lawprofessors.typepad.com/files/35374462-prop-8-ruling-final.pdf> (accessed 2/28/12)

Reinhard, David. An attack on the secret ballot. <http://www.freerepublic.com/focus/f-news/2090383> (accessed 11/4/2008)

Richard Dawkins Foundation for Reason and Science. “Gay marriage outlawed in California.” <http://richarddawkins.net/articles/3308> (accessed 2/23/2012)

San Francisco Daily Evening Bulletin, March 25-26, 1891; *San Francisco Alta California*, January 15, 1891.

“What role for Democratic 'super-delegates'?” April 26, 2007. <http://www.msnbc.com/id/18277678/>. (accessed 2/24/12)

U.S. House of Representatives. “Final vote results for Roll Call 118.” <http://clerk.house.gov/evs/2007/roll118.xml> (accessed 2/24/12)

US Legal. Definitions. <http://definitions.uslegal.com/s/superdelegate/> (accessed 2/20/12)

Vital Statistics on American Politics, 2007-2008. (2008). Washington, DC: CQ Press.

Witcover, Jules. (2003). *Party of the People*. New York: Random House p. 564.

Determinants of Racial Concentration in Mountain States Counties, 2000

L. Dwight Israelsen

Utah State University

Ryan D. Israelsen

Indiana University

Abstract

Studies of residential segregation have concentrated on racial/ethnic segregation in neighborhoods, standard metropolitan areas, counties, and states, looking at socioeconomic status and other variables to explain differences in segregation measures across groups and over time. Commonly used segregation indices measure segregation of a particular race or ancestral group in an area by examining the distribution of that group across sub-areas. The current study utilizes a new measure of segregation, the "racial concentration ratio" or RCR. The RCR is a one-parameter measure of the racial (ethnic, ancestral) composition of an area (county, state) relative to the racial composition of a larger geographical area—in this study, the Mountain States as a whole (Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming)—and is similar to a Gini coefficient. RCRs are based on the percentages of county (state) population in six groups: Non-Hispanic

Whites, Hispanics, Blacks, Asians (Asians/Pacific Islanders), American Indians (Indians and Alaskan Natives), and All Other. First, the percentage of each group in a county or state is scaled by the overall regional (Mountain States) percentage of that group. Then, RCRs are calculated, treating these cumulative scaled percentages as "cumulative income" is treated in the income Gini coefficient calculation (i.e., on the y axis of the Lorenz curve) and treating the fractions 1/6, 2/6, 3/6, 4/6, 5/6, and 6/6 as "cumulative population" is treated in the income Gini coefficient calculation (i.e., on the x axis of the Lorenz curve). Hence, if the racial composition of an area is exactly like that of the region as a whole, the RCR is zero, and as the racial composition increasingly diverges from that of the region, the RCR grows closer to one. We calculated RCRs for all Mountain states and counties. We found that statistically significant variables associated with higher county RCRs include net births of Hispanics, net births of "Others," percent speaking a language other than English at home, percent rural farm dwellers, unemployment rate, and percent of eligible voters who voted. Statistically significant variables associated with lower county RCRs include net in-migration of Blacks, net births of Blacks, population churning rate, rent, and percent of the population urban. Idaho, Montana, Utah, and Wyoming have more racial concentration and Nevada has less racial concentration than would have been predicted by the model, relative to Arizona, the control state.

I. Introduction

Studies of residential segregation have concentrated on racial/ethnic segregation in neighborhoods, standard metropolitan areas, counties, and states, looking at socioeconomic status and other variables to explain differences in segregation measures across groups and over time.¹⁻⁸ Most studies have measured segregation using census tracts or blocks and have considered segregation by neighborhoods or Metropolitan Statistical Area. Only a few studies have examined segregation at the county level, and in the most recent of those the authors have utilized a new measure of segregation, the "racial concentration ratio (RCR)."⁹⁻¹¹ Those studies identify racial concentration ratios for all U.S. states and counties and utilize large-model regression analysis to test the importance of various economic, social, education, demographic, political, and location variables associated with counties in affecting the degree of racial concentration for 2000. The current study applies the same methodology to the Mountain States region (Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Nevada).

Racial concentration ratios were calculated for each Mountain county and state, and econometric analysis was used to examine the importance of various characteristics associated with Mountain States counties in determining the degree of racial concentration in 2000.

II. Racial Concentration Ratio

Commonly used segregation indices measure segregation of a particular race or ancestral group in an area by examining the distribution of that group across sub-areas. Historically, the two most widely used measures of segregation have been the dissimilarity index and the isolation index. The dissimilarity index, put simply, is a measure of how many people in a county would have to change areas of residence to make the distribution of races even throughout the county. The dissimilarity index can be defined as

$$\frac{1}{2} \text{SUM } |(race_i / race_{\text{total}}) - (nonrace_i / nonrace_{\text{total}})|,$$

where *race* is a particular racial or ethnic group; *race_i* is the number of *race* in the *i*th census tract in the county; *race_{total}* is the number of *race* in county; *nonrace_i* is the number of individuals in the census tract not of race/ethnicity *race*; and *nonrace_{total}* is the number of individuals in the county not of race/ethnicity *race*. The dissimilarity index can range from 0, when individuals of a certain race/ethnicity are evenly distributed throughout all tracts in the county, to 1, when all of the individuals in question are living in one tract and not one individual from another race/ethnicity is living in that tract.

A second measure of segregation is the isolation index, which is defined as

$$\frac{\{\text{SUM}[(race_i/race_{\text{total}})*(race_i/persons_i)] - (race_{\text{total}}/persons_{\text{total}})\}}{\text{MIN}[(race_{\text{total}}/persons_i), 1] - (race_{\text{total}}/persons_{\text{total}})}$$

where *race_i* and *race_{total}*, are defined as above, *persons_i* is the total number of people in the tract, and *persons_{total}* is the total number of people in the county. This measure provides an estimate of the amount of isolation of the group *race* for each county in the study, corrected for population size of the group. Both of these indices are very well explained by Cutler et al.¹²⁻¹³

Echenique and Fryer¹⁴ pointed out various weaknesses found in the dissimilarity and isolation indices and proposed the mathematically intense Spectral Segregation Index (SSI), which overcomes these problems. They present correlations of the SSI with existing measures of segregation, including dissimilarity (0.42) and isolation (0.92). Because of the high correlation between the SSI and the isolation index, it ap-

pears that the isolation index is an acceptable proxy for the computationally difficult SSI.

The current study uses a new measure of segregation, the RCR. The RCR is a one-parameter measure of the racial (ethnic, ancestral) composition of a sub-area (county, state) relative to the racial composition of the larger area (state, region, United States), and is similar to a Gini coefficient. Unlike the most commonly used segregation measures, the RCR is based on multiple racial/ancestry groups, and unlike a standard Gini coefficient, the RCR places the same "population" weight on each group. RCRs are based on the percentages of county (state) population in six groups: Non-Hispanic Whites, Hispanics, Blacks, Asians (Asians/Pacific Islanders), American Indians (Indians and Alaskan Natives), and All Other. First, the percentage of each group in a county or state is scaled by the overall regional (Mountain States) percentage of that group. Then, RCRs are calculated treating these cumulative scaled percentages as "cumulative income" is treated in the income Gini coefficient calculation (i.e., on the y axis of the Lorenz curve) and treating the fractions 1/6, 2/6, 3/6, 4/6, 5/6, and 6/6 as "cumulative population" is treated in the income Gini coefficient calculation (i.e., on the x axis of the Lorenz curve). Hence, if the racial composition of an area is exactly like that of the entire Mountain States region, the RCR is zero, and as the racial composition increasingly diverges from that of the Mountain States, the RCR grows closer to one. Because there are only six groups, and the RCR is calculated from the Lorenz Curve using trapezoids, the maximum RCR is 5/6, or 0.833. RCRs were calculated for each Mountain state and county based on the scaling of their racial/ethnic groups by the overall Mountain States region percentages. Table 1 shows the 31 least-concentrated Mountain States counties and the 31 most-concentrated counties, along with their RCRs for 2000, and Table 2 ranks the Mountain States based on their RCRs in 2000.

In 2000, Mountain States county RCRs range from 0.104 in Carson County, Nevada, to 0.803 in Apache County, Arizona, with a mean of 0.476. RCRs among the Mountain States range from 0.111 in Arizona to 0.492 in Montana, with an average of 0.297. Each state except Montana is represented among the 31 least-concentrated counties, and each state except Nevada is represented among the 31 most-concentrated counties. Many of the most-concentrated counties consist wholly or in part of Indian reservations, with large RCRs reflecting overrepresentation of Indians relative to the Mountain States as a whole, while some, such as Rich and Bear Lake Counties in Utah, have high RCRs because of overrepresentation of whites. Other Mountain States counties with high RCRs have overrepresentation of Hispanics.

Table 1. Mountain States county racial concentration ratio ranking (2000)

Least-concentrated counties			Most-concentrated counties		
County	State	Ratio	County	State	Ratio
Carson	NV	0.1040	Apache	AZ	0.8028
Pima	AZ	0.1046	Glacier	MT	0.7922
Maricopa	AZ	0.1435	San Juan	UT	0.7840
Otero	NM	0.1893	Big Horn	MT	0.7807
Bernalillo	NM	0.1990	McKinley	NM	0.7796
Churchill	NV	0.2097	Roosevelt	MT	0.7762
White Pine	NV	0.2151	Blaine	MT	0.7752
Pershing	NV	0.2320	Navajo	AZ	0.7353
Elmore	ID	0.2344	Petroleum	MT	0.7306
Cochise	AZ	0.2380	Rosebud	MT	0.7255
Adams	CO	0.2397	Mora	NM	0.7190
Weber	UT	0.2472	Liberty	MT	0.7138
Laramie	WY	0.2531	San Juan	NM	0.7058
Tooele	UT	0.2568	Chouteau	MT	0.7023
Humboldt	NV	0.2667	Cibola	NM	0.6971
Washoe	NV	0.2674	Garfield	MT	0.6968
Mohave	AZ	0.2800	Santa Cruz	AZ	0.6893
Elko	NV	0.2844	Golden Valley	MT	0.6871
Carbon	WY	0.2853	Judith Basin	MT	0.6797
Albany	WY	0.2871	Lake	MT	0.6788
Bent	CO	0.3041	Pondera	MT	0.6743
Chaffee	CO	0.3057	Carter	MT	0.6662
Catron	NM	0.3083	Fremont	WY	0.6615
Lyon	NV	0.3090	Rich	UT	0.6609
Owyhee	ID	0.3146	Hill	MT	0.6560
Nye	NV	0.3173	Morgan	UT	0.6507
Bannock	ID	0.3196	Prairie	MT	0.6458
Huerfano	CO	0.3199	Hidalgo	NM	0.6453
Curry	NM	0.3201	Bear Lake	ID	0.6425
Pinal	AZ	0.3203	Conejos	CO	0.6354
Salt Lake	UT	0.3205	Powder River	MT	0.6328

Table 2. Mountain States racial concentration ratio ranking, 2000	
State	Ratio
AZ	0.1106
CO	0.1918
UT	0.2816
NV	0.2925
WY	0.3134
ID	0.3284
NM	0.3636
MT	0.4915

Figures 1 and 2 show county and state RCR maps, with lighter colors indicating lower levels of racial/ethnic concentration and darker colors indicating higher levels of concentration.

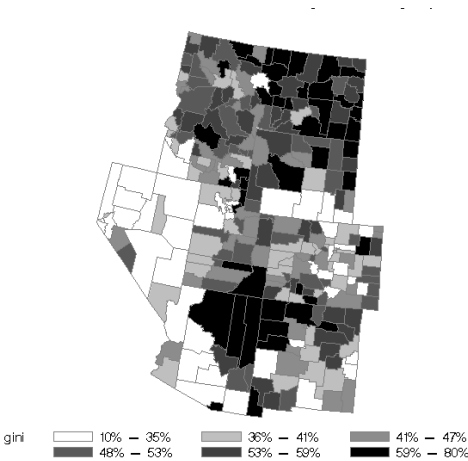


Figure 1. Racial concentration ratio by county (2000)

The state RCR map shows a reduction in the degree of racial concentration in the Mountain States from north to south, with New Mexico in the far southeast of the region being an exception. A closer look at specific racial/ethnic groups below reveals the heavy concentration of Hispanics and American Indians in New Mexico and the underrepresentation of the other groups that contribute to New Mexico’s high

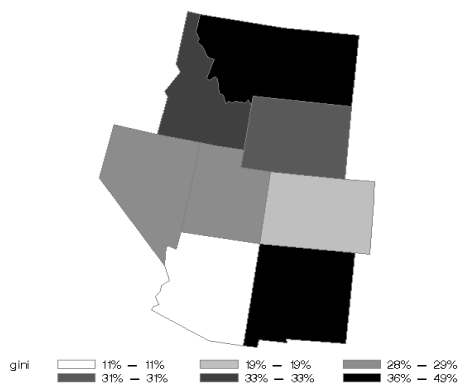


Figure 2. Racial concentration ratio by state (2000)

RCR. The county RCR map reveals a more accurate picture of the two areas of greatest racial/ethnic concentration, one centered in the Four Corners region and the other extending from northeastern Montana south and west to include all but the southern counties of Wyoming and the southwestern counties of Idaho, and dipping into northeastern Utah. Counties containing or consisting of Indian reservations are among those with the highest RCRs. Areas of low racial/ethnic concentration extend through central Colorado north to the tier of counties in southern Wyoming, then west to the border counties. The low-RCR region begins again in western Utah, extends through Nevada, and encompasses most of southwestern Arizona and southeastern New Mexico. The importance of measuring racial concentration at the county, rather than state, level is seen by looking at Arizona. Whereas almost the entire northeast half of Arizona contains counties with very high RCRs, the large majority of Arizona's population lies in the southwest half where almost all counties have low RCRs; hence, the state as a whole has the lowest RCR of all Mountain States. Inversely, Nevada, in which all but four counties have RCRs in the smallest category, has a higher RCR than Arizona because Nevada's two most populous counties have moderately high levels of racial concentration.

III. Distribution of Racial/Ethnic Groups

Figures 3–14 show the relative distribution of each racial/ethnic group in Mountain states and counties. The percentage of each racial/ethnic group in a state or county is scaled by the overall percentage of that group in the Mountain States. Hence, if a particular racial or

ethnic group makes up the same percentage of the overall population in a county as it does in the Mountain States, that group's scaled percentage in the county would be 1.0. If the group is "under-represented" in the county relative to the Mountain States, its scaled percentage would be less than 1, and if it is "over-represented" in the county, its scaled percentage would be greater than 1.

There are several interesting facts about the relative percentage of racial/ethnic groups that are evident from the maps. First, it is not surprising that the largest group, Whites, is also the least-concentrated, as indicated by scaled percentages close to 1.0. It is somewhat surprising that the "Other" category is the second-least concentrated group. Though few in numbers, members of "other" racial and ethnic groups are distributed quite evenly across the populations in Mountain states and counties.

The most-concentrated Mountain States racial/ethnic group is Blacks, with very low relative percentages in the northern and central parts of the Mountain States, and high relative percentages in the border areas of eastern Colorado, southern New Mexico and Arizona, and in four Nevada counties. The percentage distribution of Hispanics is somewhat more equal than the relative distribution of Blacks, but it follows the same pattern, generally increasing from north to south across the region and being very high in New Mexico, southern Colorado, and southern Arizona.

Asians are concentrated in urban areas, especially around Denver, Salt Lake City, Reno, and Las Vegas. American Indians are heavily concentrated in counties containing reservations and, hence, in Montana, Arizona, and New Mexico.

Finally, a comparison of county and state maps shows that the demographics of Clark County dominate the overall demographic profile of Nevada.

IV. Factors Creating Differences in RCRs

Differences in RCRs across counties and states must be the result of differential demographic forces operating over time, specifically, natural population change by race/ethnic group, in-migration, out-migration, and net immigration from abroad by racial/ethnic group. Differences in RCRs develop in response to differences in demographic, economic, social, education, political, and other factors that

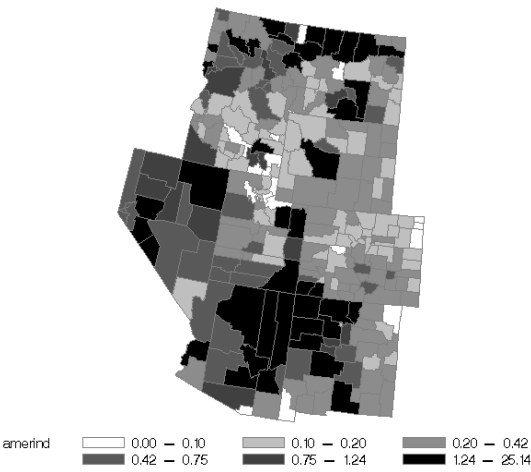


Figure 3. Relative ratio of Amerind by county (2000)

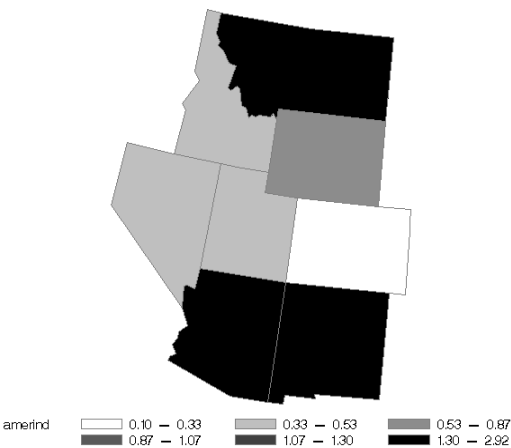


Figure 4. Relative ratio of Amerind by state (2000)

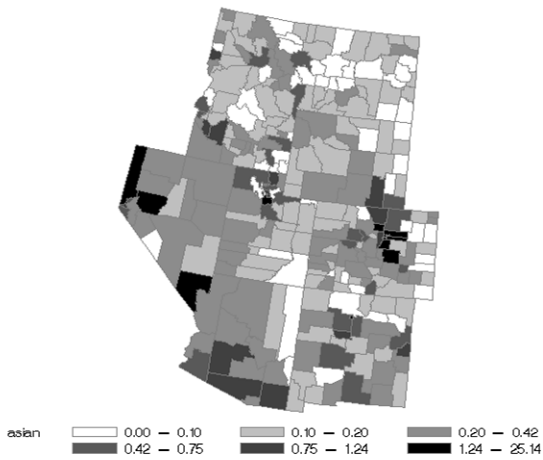


Figure 5. Relative ratio of Asian by county (2000)

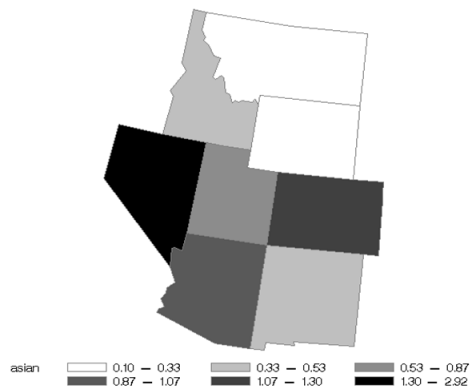


Figure 6. Relative ratio of Asian by state (2000)

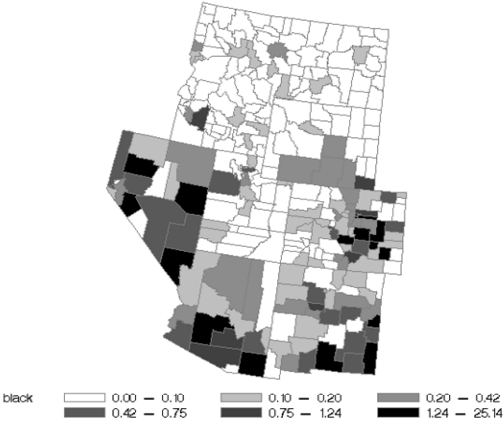


Figure 7. Relative ratio of Black by county (2000)

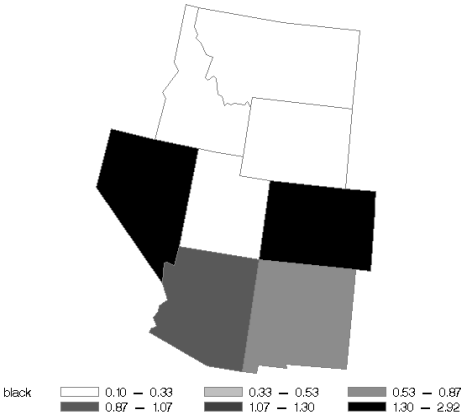


Figure 8. Relative ratio of Black by state (2000)

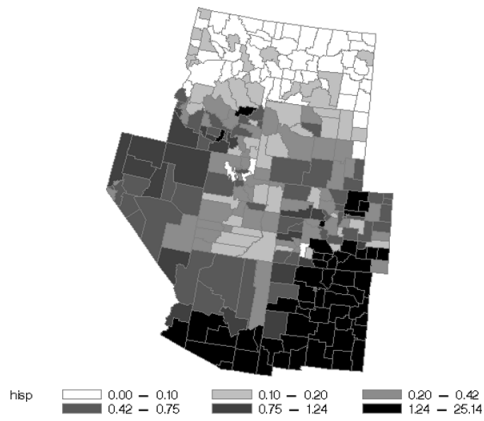


Figure 9. Relative ratio of Hispanic by county (2000)

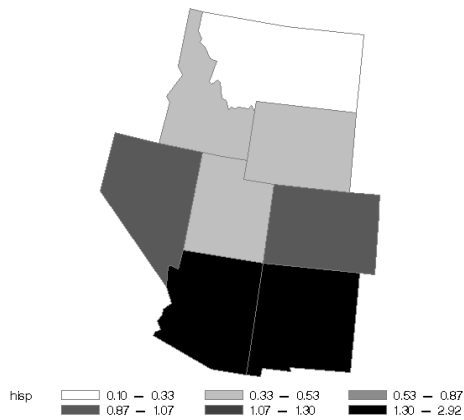


Figure 10. Relative ratio of Hispanic by state (2000)

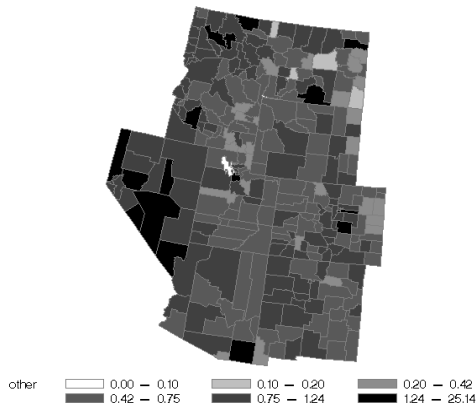


Figure 11. Relative ratio of Other by county (2000)

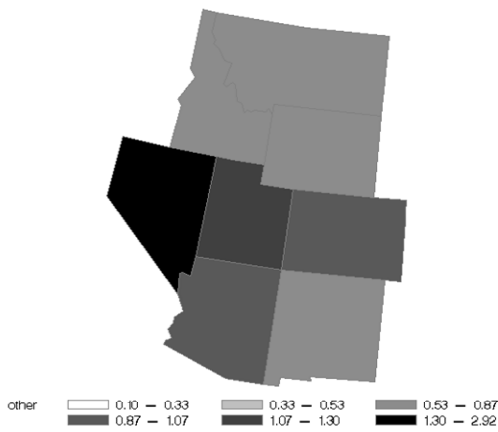


Figure 12. Relative ratio of Other by state (2000)

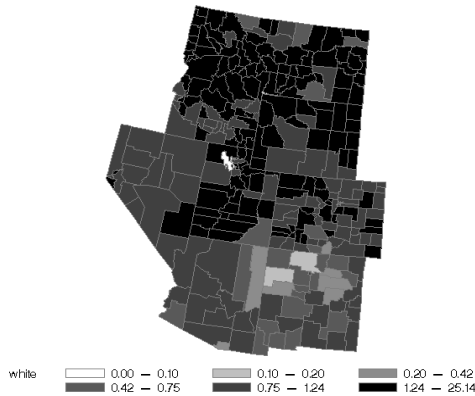


Figure 13. Relative ratio of White by county (2000)

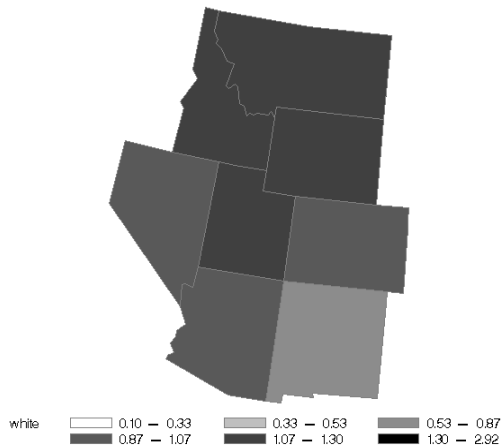


Figure 14. Relative ratio of White by state (2000)

help determine how changes in population are distributed among counties and states within the Mountain States region. An econometric model is used to identify the importance of each of these factors in influencing racial concentration within counties, using county RCR scaled by the overall Mountain States population as the dependent variable in the model.

Independent Variables

Independent variables in the model are characteristics of counties that can be grouped into demographic, economic, social, education, and political factors.

Demographic variables

Demographic variables in the model include net in-migration of Whites, Blacks, Hispanics, and Others (1995–2000 as percent of population); net births rates of Whites, Blacks, Hispanics, and Others (percent of population); percent speaking a language other than English at home; percent foreign-born; average age; average household size; percent urban dwellers; percent rural farm dwellers; and population per square mile. Population churning rate (pcr), which is the sum of gross in-migration and gross out-migration, is included as an alternative to racial/ethnic group in-migration rates.

Economic variables

Economic variables include percent below the poverty line, real per capita income, unemployment rate, average real mortgage payment, average real rent, tax burden, government revenue per capita, government debt per capita, local government employment per capita, and property tax per capita.

Social variables

Social variables include percent of households married, violent crime rate, and crimes per 100,000.

Education variables

Two education variables are used in the model, percent of population with at least 12 years' schooling and percent of population with higher education.

Political variables

Political variables include percent of eligible voters who actually voted in the 2000 Presidential election and percent of voters who voted for the Democrat candidate in the 2000 election.

State dummy variables

State dummy variables are used for 7 Mountain states, with Arizona the control.

In general, we did not specify expected signs for the estimated coefficients.

Models

Four versions of the model of the determinants of county RCR were formulated and tested econometrically. Model 1 includes all independent variables except pcr. Model 2 is the same as Model 1 except for the exclusion of mortgage, rent, and violent crimes, which are highly multicollinear with similar variables. Model 3 is the same as Model 1 except for the replacement of net in-migration variables with pcr. Model 4 is the same as Model 2, except for the replacement of net in-migration variables with pcr. Models 1 and 3 were tested as the two basic alternatives, differing only in the migration variable(s) used. Because of very high multicollinearity between real per capita income, mortgage, and rent and between violent crimes and crime index (as indicated by variance inflation factors), Models 2 and 4 were also tested to check the robustness of the results.

Results

Table 3 shows the statistically significant coefficients in the four regressions, along with the signs and levels of statistical significance of the estimated coefficients.¹

Degrees of freedom in the models range from 241 (Model 1) to 247 (Model 4). The F-statistic in all models is significant at 0.0001 and adjusted R^2 ranges from 0.655 (Model 1) to 0.664 (Model 3).

Statistically significant variables associated with lower county RCRs include net in-migration by Blacks, net births of Blacks, population churning rate, rent, and percent of the population urban. These results are consistent with a pattern of migration of Blacks from areas in which they are relatively over-represented into areas in which they are relatively under-represented. In addition, net births of Blacks appear to be relatively higher in areas in which they are under-represented and lower in areas in which they are over-represented. The negative estimated coefficient of pcr indicates that, other things being equal, higher population mobility leads to lower racial/ethnic/ancestral concentration in Mountain States counties relative to the racial/ethnic/ancestral composition of the Mountain States as a whole. The process of urbanization in the Mountain States also contributes to lower RCRs.

¹ Detailed results are available from the author, disraelen@econ.usu.edu.

Table 3. Statistically significant coefficients: signs and significance levels

Variable	Model 1	Model 2	Model 3	Model 4
Net in-migration – Blacks	— ^{**}	— ^{**}	n.a.	n.a.
Population churning rate	n.a.	n.a.	— ^{**}	— ^{**}
Net births – Blacks	— [*]	— [*]	— [*]	— [*]
Net births – Hispanics	+ [*]	+ [*]	+ [*]	+ [*]
Net births – Other	+ ^{***}	+ ^{***}	+ ^{***}	+ ^{***}
Language	+ ^{**}	+ ^{**}	+ ^{**}	+ ^{**}
Unemployment	+ [*]	+ [*]	+ ^{**}	+ ^{**}
Rent	— ^{n.s.}	n.a.	— [*]	n.a.
% Urban	— ^{**}	— ^{**}	— ^{***}	— ^{***}
% Rural Farm	+ ^{**}	+ ^{**}	+ [*]	+ [*]
%Voted	+ [*]	+ [*]	+ ^{**}	+ ^{**}
Idaho	+ [*]	+ [*]	+ [*]	+ [*]
Montana	+ ^{**}	+ ^{**}	+ ^{**}	+ ^{**}
Nevada	— ^{**}	— ^{**}	— ^{**}	— ^{**}
Utah	+ [*]	+ [*]	+ ^{n.s.}	+ ^{n.s.}
Wyoming	+ [*]	+ ^{**}	+ [*]	+ ^{**}

*significant at 0.05; **significant at 0.01; ***significant at 0.0001; n.s. = not significant at 0.05; n.a.= variable not in model

Statistically significant variables associated with higher county RCRs are net births of Hispanics, net births of Others, percent of county population speaking a language other than English at home, unemployment rate, percent of county population classified as rural farm, and percent of population who voted. Unlike net births of Blacks, net births of Hispanics and Others are associated with higher county RCRs, *ceteris paribus*. The positive sign on these estimated coefficients is consistent with net births of Hispanics and Others being higher in counties in which those racial/ethnic groups are over-represented and lower in counties in which they are under-represented. The positive estimated coefficients for the percent of the population speaking a language other than English at home suggests that failure of groups to assimilate into the English-speaking culture leads to greater racial/ethnic concentration. The positive coefficient for unemployment indicates that higher unemployment rates are associated with areas of greater racial/ethnic concentration, *ceteris paribus*. The results also indicate that counties with higher percentages of farmers are likely to also have higher RCRs, other things equal. Finally, the estimated coefficient on percent of registered voters in a county who actually voted is positively associated

with higher levels of racial/ethnic concentration. A natural interpretation of this result is that racially/ethnically homogeneous populations create environments that encourage greater political involvement, rather than the alternative causal relationship that political involvement leads to more homogeneous (racially concentrated) communities.

The signs of the coefficients estimated for state dummy variables reveal that, relative to Arizona counties, Idaho, Montana, Utah, and Wyoming counties have higher RCRs than would have been expected based on the other variables in the model, and Nevada counties have lower RCRs than would have been expected.

The consistency of signs and significance of the estimated coefficients in the four versions of the model indicates that the basic model is robust to specification variation and gives added confidence to the results.

V. Conclusions

Racial and ethnic concentration, as measured by the RCR, varies widely across the eight Mountain States and among counties in each state. At the state level, there is a general reduction in the degree of racial concentration in Mountain States from north to south, with New Mexico in the far southeast of the region being an exception. In New Mexico, the heavy concentration of Hispanics and American Indians and the under-representation of the other groups contribute to New Mexico's high RCR. At the county level, one of the two areas of greatest racial/ethnic concentration is centered in the Four Corners region, and the other extends from northeastern Montana south and west to include all but the southern counties of Wyoming and the southwestern counties of Idaho, and dips into northeastern Utah. Counties containing or consisting of Indian reservations are among those with the highest RCRs. Areas of low racial/ethnic concentration extend through central Colorado north to the tier of counties in southern Wyoming, then west to the border counties. The low-RCR region begins again in western Utah, extends through Nevada, and encompasses most of southwestern Arizona and southeastern New Mexico.

Among Mountain States racial and ethnic groups, the least concentrated, as indicated by a relative percentage of state or county population (scaled by the corresponding percentage of the entire Mountain States population) close to 1.0, is Whites, which is perhaps not surprising since it is the largest racial/ethnic group. Somewhat surprisingly, the next least-concentrated group is Others, which are distributed relatively uniformly across Mountain states and counties despite being few in numbers. The most-concentrated group is Blacks, with very low rela-

tive percentages in the northern and central parts of the Mountain States and high relative percentages in the border areas of eastern Colorado, southern New Mexico and Arizona, and in four Nevada counties. The percentage distribution of Hispanics is somewhat more equal than the relative distribution of Blacks, but it follows the same pattern, generally increasing from north to south across the region and being very high in New Mexico, southern Colorado, and southern Arizona. Asians are concentrated in urban areas, especially around Denver, Salt Lake City, Reno, and Las Vegas, and American Indians are heavily concentrated in counties containing reservations, namely, in the states of Montana, Arizona, and New Mexico.

Econometric analysis of the relationship between various demographic, economic, social, education, and political factors and RCRs indicates that net in-migration by Blacks, net births of Blacks, pcr, rent, and percent of the population that is urban are statistically significant determinants of lower RCRs and that net births of Hispanics, net births of Others, percent of county population speaking a language other than English at home, unemployment rate, percent of county population classified as rural farm, and percent of population who voted are statistically significant determinants of higher RCRs.

These results will be useful to those who wish to formulate public policy designed to change racial/ethnic concentration rates in Rocky Mountain states and counties. An obvious example would be the use of assimilation programs to reduce racial and ethnic concentration. Similarly, programs that are successful in reducing unemployment would also be expected to reduce RCRs.

Notes

¹ Giles, Michael, et al. (June 1976) Determinants of Resegregation: Compliance Rejection. Behavior and Policy Alternatives

² Crowder, Kyle. The Racial Context of White Mobility: An Individual-Level Assessment of the White Flight Hypothesis. Social Science Research, 29, 223-257 (2000)

³ Iceland, John, Daniel H. Weinberg, and Erika Steinmetz, U.S. Census Bureau, Series CENSR-3, Racial and Ethnic Residential Segregation in the United States: 1980-2000, U.S. Government Printing Office, Washington, DC, 2002

⁴ Friedman, Samantha. Do declines in residential segregation mean

stable neighborhood racial integration in metropolitan America? A research note. Social Science Research 37 (2008) 920–933

⁵ Krysan, Maria. Does race matter in the search for housing? An exploratory study of search strategies, experiences, and locations. Social Science Research 37 (2008) 581–603

⁶ Douglas S. Massey and Mary J. Fischer. Does Rising Income Bring Integration? New Results for Blacks, Hispanics, and Asians in 1990. Social Science Research 28, 316–326 (1999)

⁷ John Iceland, Cicely Sharpe, and Erika Steinmetz. Class differences in African American residential patterns in US metropolitan areas: 1990–2000. Social Science Research 34 (2005) 252–266

⁸ David M. Cutler, Edward L. Glaeser, Jacob L. Vigdor. When are ghettos bad? Lessons from immigrant segregation in the United States. Journal of Urban Economics 63 (2008) 759–774.

⁹ L. Dwight Israelsen, Ryan D. Israelsen, and William J. Israelsen. Is the United States Becoming Regregated? Utah State University ERI Study Paper, June 2008

¹⁰ L. Dwight Israelsen and Ryan D. Israelsen. Racial Concentration in U.S. Counties, 2000. Utah State Univ. ERI Study Paper, June 2009

¹¹ L. Dwight Israelsen and Ryan D. Israelsen. Determinants of Racial Concentration in U.S. Counties, 2000. Utah State University ERI Study Paper, June 2010.

¹² Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. *The Rise and Decline of the American Ghetto*. Journal of Political Economy, 1999, vol. 107, no. 3

¹³ Cutler, Glaeser, Vigdor. *Is the Melting Pot Still Hot? Explaining the Resurgence of Immigrant Segregation*. (2004) <http://trinity.aas.duke.edu/~jvigdor/cgv2004a.pdf>.

¹⁴ Echenique, Federico and Roland G. Fryer, Jr. *A measure of segregation based on social interactions*. The Quarterly Journal of Economics May 2007, vol. 72, issue 2, pp. 441–485.

Antigay Behaviors Among Young Adults: Prevalence, Patterns, and Motivators in a Noncriminal Population (Replicated)

CoCo James, Chad Morrow, and Spencer Blake
Salt Lake Community College

Abstract

This 2011 study, undertaken in Salt Lake City, Utah, was a replication of work done by Dr. Karen Franklin in the San Francisco Bay area in 1998. Both studies focused on rates of antigay attitudes and behaviors among community college students. A smaller percentage of respondents in Salt Lake City reported participation in incidents of antigay violence, but there were similar rates of name calling in both studies. Differences were discovered in demographic data (particularly rates of religious identity and racial demographics) and in the relationships between assailants and targeted homosexuals.

Introduction

The brutal 1998 murder of openly gay University of Wyoming student Matthew Shepard shocked American society and brought antigay violence to the forefront of national dialogue. Dr. Karen Franklin, a clinical and forensic psychologist in the San Francisco Bay area, was interested in the political and social climate of her region in the months after the Shepard murder. She developed a survey and administered it at six community college campuses in the San Francisco Bay area. In her words, "The dual purposes of this study were to explore the self-described motivations of individuals who have committed antigay behaviors and to obtain preliminary rates for specific antigay behaviors among a noncriminal young adult population." (Franklin 2000).

In 2010, a rash of suicides by young gay men again brought national attention to the treatment of homosexuals in American society. These suicides prompted the question of how attitudes and behaviors had changed in the 12 years between these events and the Shepard murder, as well as how different populations would approach the same issue.

Overview

The findings from a single study are far less convincing than findings from a series of related studies. Replication often reveals findings that are slightly different from the original work. Other times, replicated work shows commonality. This allows the research topic to gain creditability with each replication. The purpose of this study was to replicate the research of Dr. Karen Franklin. Dr. Franklin conducted her study in 1998 and published her findings in a paper entitled *Antigay Behaviors Among Young Adults: Prevalence, Patterns, and Motivators in a Noncriminal Population* (Franklin 2000). This was "the first empirical research into the prevalence rates of and motivations for antigay harassment and violence by noncriminal young adults." (Franklin 2000). Her study was replicated in Salt Lake City, Utah, in 2011 at Salt Lake Community College (SLCC).

Literature Review

Dr. Franklin formulated many of her questions in response to the current research of the day. Rates of victimization due to sexual orientation were reported by the National Gay and Lesbian Task Force in *Anti-gay violence, victimization and defamation in 1989*. The pervasiveness of antigay bias was established in a study by Alan S. Yang, *Trends: Attitudes toward homosexuality* (Yang 1997). Dr. Franklin also

used the *Attitudes Toward Lesbians and Gay Men Scale* (1994) created by Dr. Gregory M. Herek, Professor of Psychology at University of California at Davis, an internationally recognized expert on prejudice towards homosexuals, including hate crimes and antigay violence. Gallup's annual Values and Beliefs poll¹ provided some items for Dr. Franklin's questioning.

In the years between the San Francisco study and this study, reports of violence against gays were on the rise. Every year between 1998 and 2011, the Federal Bureau of Investigation (2011) has reported an increase in violence motivated by anti-gay bias; however, according to Gallup's annual Values and Beliefs poll, national rates of acceptance of homosexuality have also continued to rise. In 2008, Americans were evenly divided on the issue, with 48% considering homosexual behavior morally acceptable and 48% considering it morally wrong (Gallup Polls 2008). Just two years later, acceptance of homosexual behavior crossed into the majority, with 50% of Americans considering the behavior morally acceptable compared with 43% considering the behavior morally wrong (Gallup Polls 2010). This trend continues to be stratified religiously. It is generally predicted that there is a negative correlation between religiosity and acceptance of homosexual behavior (Green et al. 2010). Because of Utah's large population of members of the Church of Jesus Christ of Latter-Day Saints (LDS), it was hoped this study would address a gap in knowledge about how people in a predominantly LDS population would feel and behave toward homosexuals.

Research Procedures

The intention of this study was to mirror the San Francisco study as closely as possible. Because of this, Dr. Franklin's procedures were the basis for the procedures of the Salt Lake City study. An exception was that Salt Lake City respondents were asked to take their surveys home and return them during the next class period whereas Dr. Franklin asked her respondents to complete their surveys during class time. This choice was made in Salt Lake City because it was undesirable to take class time away from the students and teachers who participated in the study. Data were analyzed using SPSS.

¹ Gallup Polls have been measuring public opinion since 1958. Each May, Gallup measures attitudes toward homosexuality in the annual Values and Beliefs poll.

Sampling Procedures

In keeping with Dr. Franklin's sampling procedures, the Salt Lake City survey was not offered to students attending four-year universities. In both regions it was believed that community college students would be more racially and economically diverse than university students. Multiple community college campuses were sampled in both San Francisco and Salt Lake City to diversify as much as possible. General education classes were chosen to eliminate any selection bias. The Salt Lake City sample size was 420 individuals; the sample size in San Francisco was 489 individuals. Because of enrollment numbers at each school, the samples were similarly representative. Questions in the survey solicited information on opinions about homosexuality (see Appendix) as well as actions toward homosexual individuals.

Findings

Definitions

There are a few instances of specific terminology in this work. In the interest of clarification, the following definitions are applicable:

Antigay behavior: Behavior that is in opposition or is hostile to homosexuals or homosexual social reform, measured in this study by name calling and physical aggression.

Assailants: Those respondents who reported participating in physical aggression toward homosexuals.

Name-callers: Those respondents who reported calling homosexuals derogatory names. Name-callers did not participate in any physical threatening of homosexuals.

Non-assailants: Those respondents who did not participate in any antigay behavior at all, either name calling or physical aggression.

Demographics

The Salt Lake City sample was composed of 43% men and 57% women, with a mean age of 25 and an age range of 18 to 59. Reported ethnicity was 77% White, 8% Latino/Hispanic, 6% Asian, 2% African American, and 1% Native American. Another 6% reported being of mixed race or other. The sample's gender and racial composition was an accurate reflection of the student body at SLCC for the 2010–2011 school year. Half of the sample (51%) reported having a father who either had some college education or had completed a college degree. Similar numbers of respondents (54%) indicated their mother had a similar educational level. One in ten respondents was an immigrant to

the United States, with 70% of those having immigrated at or before the age of 17 years.

A majority of the respondents (64%) were single, 17% were married and living with their spouse, and another 9% were living with a partner but were not legally married. About 57% of the sample belonged to an organized religion, and 42% reported going to church once a week or more. The sample was politically diverse, with 37% identifying with a conservative political outlook, 30% identifying with a “middle of the road” political outlook, and 29% identifying with a liberal political outlook. In terms of party affiliation, 34% were Republican, 19% were Democrat, 25% were Independents, and 22% identified as something else altogether.

Comparisons

With this replication, two noteworthy distinctions should be emphasized. These distinctions were cultural in nature, not systemic with the research. First, the Salt Lake City population sample came from a “churched” county whereas the San Francisco population did not. Second, the Salt Lake City sample was not as ethnically diverse. Both of these conditions are inherently demographic. A sample taken from Utah and one taken from California would be slightly different no matter which random technique is used.

The most significant difference between the Salt Lake City study and the San Francisco study was in the reported rates of religious affiliation. Nearly all of Franklin’s respondents (97%) identified with an organized religion. The Salt Lake City figure was 56.5%. This was most surprising given the religiosity of each respective county during the time the data was collected, not to mention the perceived cultural and religious identity of each location. The difference in self-reported religious affiliation between the areas was significant enough that a mistake appeared to have been made. The Salt Lake City data were revisited and found to be accurate. Methods of data analysis were also reconsidered and found to be consistent between the two studies. Dr. Franklin also revisited the data in the San Francisco study and confirmed its accuracy. No explanation appears to be inherent in the data from either study. This curious finding begs for further exploration, however, as it was not the focus of this study, no analysis of this phenomenon will be provided here.

In the Salt Lake City study, 22 respondents (approximately 5% of the sample) reported having physically assaulted or threatened people they believed to be homosexual. Another 96 respondents (22.9%) reported calling homosexuals by insulting names but not physically as-

saulting them. In the San Francisco study, 49 respondents (10%) reported assailant behavior; however, rates of name-calling remained nearly consistent, with the San Francisco study reporting 23.5% compared with the Salt Lake City study at 22.9%.

Of the 302 respondents in the Salt Lake City study (72%) who denied ever having harassed homosexuals, 201 respondents (48%) reported having friends who had. Twenty-one percent of those 302 respondents reported having witnessed incidents of antigay behavior. Only 86 respondents (approximately 20%) in the Salt Lake City study reported never having engaged in antigay behavior, did not have friends who had engaged in antigay behavior, and had not ever witnessed an incident of antigay behavior. In the San Francisco study, 321 respondents (66.3%) denied any antigay behavior of their own, but 83 respondents (23%) reported witnessing such things.

One of the characteristics that distinguished assailants from non-assailants in both studies was gender. Males had higher rates of antigay aggression in both studies. In the Salt Lake City study, males represented 43% of the respondents and 63.8% of the perpetrators. This was a near-perfect mirror of the San Francisco study, in which males represented 43% of the respondents and 64% of the perpetrators.

This propensity toward antigay aggression in males is also evident among peer groups. Three out of four respondents in the Salt Lake City study reported having male friends who had directly called homosexuals an insulting name in comparison with 53% of their female friends. This is reasonably consistent with the San Francisco study, where 69% of respondents said their male friends had engaged in name-calling compared with just over half of their female friends.

As in the San Francisco study, the next most common antigay behavior by respondents' friends in the Salt Lake City study was participating in verbal threats. In the Salt Lake City study, 29.8% of respondents reported having male friends who had verbally threatened homosexuals. An additional 9.4% of respondents' male friends were reported as having hit, kicked, or beaten a homosexual. These percentages are similar to the San Francisco study, which reported 29% and 7%, respectively.

Of the 118 respondents in the Salt Lake City study who admitted to engaging in some kind of antigay behavior, 84 of them answered the set of questions in the survey about a specific incident. This included 64 respondents who had engaged in name-calling only and 20 who had threatened or physically assaulted homosexuals. A majority of the incidents (64%) occurred two years ago or more. This differs from the San Francisco study, wherein more than half of all incidents occurred within the previous two years. In the Salt Lake study, the respondents' me-

dian age at the time of the reported incidents was 17 years (range = 11–32) compared with a median age of 18 years (range = 10–40 years) in the San Francisco study. Most reported incidents in the Salt Lake City study (58%) occurred either at school or the workplace. This was also the most common response in the San Francisco study.

Only 12% of respondents from the Salt Lake City study reported drinking alcohol and 8% reported using drugs (marijuana being the most common) before or during the incident. This rate of alcohol and drug use is consistent with the San Francisco study, where 16% of assailants reported having consumed alcohol and 7% reported drug use before or during the incidents.

In both the Salt Lake City and the San Francisco studies, most assailants acted in groups. In the Salt Lake City study, 88% of respondents were part of a group, with 51% being with 3 or more people. Of those in groups, 43.6% were with friends, 26.7% were with classmates, coworkers, or teammates, and 9.9% were with family members. About half (47.9%) of respondents said their level of involvement in the incident was at least that of their companions, and 21.1% said they had started the incident. This is consistent with the findings of the San Francisco study in which “a broad majority of assailants were with other people during the incidents they described.” (Franklin 2000) The largest percentage (roughly 2/3) was with their friends, 20.5% were with classmates, coworkers, or teammates, and 11.5% were with family members. Most assailants in the San Francisco study reported playing at least as great a role as their companions in the incidents.

In contrast to the group nature of assailants in both the Salt Lake City and San Francisco studies, the presumed homosexuals targeted in antigay incidents were overwhelmingly more likely to be alone. In the Salt Lake City study, 70% of respondents reported the presumed homosexual was alone during the incident, and another 22% said there were only two presumed homosexuals. There was not a single incident wherein one assailant targeted more than two presumed homosexuals together. These figures were similar to those from the San Francisco study in which 2/3 of the incidents involved a single presumed homosexual, 15% involved two presumed homosexuals, and 18% involved three or more presumed homosexuals.

The targeted individuals in both studies were primarily male. The Salt Lake City study reported 86% of presumed homosexual targets were male, 6% were female, and 7% of the incidents included both male and female presumed homosexuals. The San Francisco study showed a similar pattern with 77.4% male, 13.2% female, and 9.4% male and female presumed homosexuals together.

The similarities between the studies became more diversified in terms of the relationships between assailants and targeted homosexuals. In the Salt Lake City study, 29% of the presumed homosexuals were complete strangers to the respondent, 34% were fellow classmates, 12% were a friend of a friend, 4.8% were coworkers, 3.6% were family members, and 1.2% were neighbors, and another 7% listed as "other." The San Francisco study showed different trends in that 63% of the presumed homosexuals were strangers to the respondent, 6% were fellow classmates, 10% were a friend of a friend, 5% were coworkers, 3% were family members, and 5% were neighbors.

Summary

In summary, the results of the Salt Lake City study were very similar to the results of the San Francisco study. Many of the patterns that were discovered by Dr. Franklin held true in the Salt Lake City study. Respondents separated into three behavior categories: assailants, name-callers, and non-assailants. Respondents in both studies had similar motivations for antigay behavior (in regards to assailants and name-callers) as well as reasons for restraint (in regards to non-assailants.) Antigay behavior was found to be pervasive in both areas, with the majority of respondents either having participated in or witnessing such events. In both studies, males were more likely to be aggressors as well as targets in antigay incidents. Alcohol and drug use had occurred in less than one out of five cases in both areas. Antigay behavior was commonly undertaken in groups and targeted homosexuals were likely to be alone.

Not all patterns held true. Antigay incidents in the San Francisco study had typically occurred in the two years preceding the study whereas in the Salt Lake City study, antigay incidents had typically occurred more than two years before the data was collected. Also, the relationships between assailants and their targets varied between the two areas. In the Salt Lake City study, most targets were the classmates of assailants whereas the San Francisco study indicated that most targets were complete strangers to the assailants.

Limitations of Study

No empirical investigation of social phenomena is free from limitations or qualifications. The following is a short rehearsal of some limiting factors regarding the Salt Lake City study.

The last four questions of Dr. Franklin's questionnaire were inadvertently excluded from the questionnaire administered at SLCC because of an undiscovered printing error at the time. These questions

included information about sexual orientation, levels of urbanization, practices of social drinking, and membership in all-male or all-female social clubs.

Dr. Franklin cited limitations within her study due to the specific cultural character of the San Francisco Bay area. She believed that her results may not be generalizable because of San Francisco's reputation as a "bastion of progressive politics and social tolerance." (Franklin 2000). In contrast, Salt Lake City has a reputation of being politically conservative and socially rigid. Both communities have distinct cultural characteristics outside of the general norms of American society that may make generalization problematic.

The total number of assailants in the Salt Lake City study was 22. Because of this small number, only tentative conclusions could be drawn. This also meant that a single respondent's survey could drastically change the outcome of any analysis of the nature of the assailant population.

Because of the stratification of the sample, neither age nor level of education could be used as a means of analysis. Most respondents in the Salt Lake City study ranged in age from 18 to 24. Additionally, every respondent except one reported "some college" as their level of education. As the survey was conducted at a community college, this was unsurprising; however, the sample does not provide the ability to compare either age or levels of education with rates of anti-gay behavior.

Future Study

Several interesting points in the data invite further examination and future research. First and foremost, the difference in self-reported religious affiliation between the two studies begs for an explanation. Some recent studies suggest that the fastest-growing category of religious affiliation in the United States is "none" (Putnam and Campbell 2010). This may account for the discrepancy in rates of reported religious affiliation but further inquiry is needed.

In each study, either moral or religious beliefs were cited as both the primary reason for restraint as well as the primary reason for aggression. This finding deserves more careful attention. The role of religion as it supports or condemns anti-gay behavior should be part of the dialogue as the nation seeks for peaceful coexistence between ideologies. Also, because many respondents did not indicate an affiliation with a particular religious faith, it would be interesting to ascertain the specifics of a moral code *not* introduced and supported through a reli-

gious organization as it pertains to Lesbian, Gay, Bisexual and Transgender issues.

Additionally, many respondents indicated their belief that name-calling was an acceptable practice as long as it was not done in the presence of a homosexual. The sociolinguistic implications of this behavior would be interesting to study, particularly in regards to the marginalization of a minority group.

Acknowledgments

First and foremost, we would like to acknowledge Dr. Karen Franklin and thank her for sharing her questionnaire, her dissertation, and her feedback on our findings. We also express gratitude to both the Salt Lake Community College professors who allowed us to come into their classes and to the students who participated in the study. Additionally, special thanks go to Karine Agajanian and Jen Chapman for their assistance with data collection as well as data entry.

Bibliography

Federal Bureau of Investigation. *Federal Bureau of Investigation Hate Crime Statistics 2010*. <http://www.fbi.gov/stats-services/crimestats> (accessed June 12, 2011).

Franklin, Karen. "Antigay Behaviors Among Young Adults: Prevalence, Patterns, and Motivators in a Noncriminal Population ." *Journal of Interpersonal Violence* , 2000: 339-362.

Gallup Polls. *Americans Evenly Divided on Morality of Homosexuality*. May 2008. <http://www.gallup.com/poll/108115/americans-evenly-divided-morality-homosexuality.aspx> (accessed June 07, 2011).

Gallup Polls. *Americans Acceptance of Gay Relations Crosses 50% Threshold*. May 2010. <http://www.gallup.com/poll/135764/americans-acceptance-gay-relations-crosses-threshold.aspx> (accessed June 07, 2011).

Gallup Polls. *Gallup Poll Values and Beliefs Survey*. May 2011. <http://www.gallup.com/search/default.aspx?q=values+and+beliefs> (accessed June 07, 2011).

Green, Mary S., Murphy, Megan J., & Blumer, Markie L. C. "Marriage

and Family Therapists' Comfort Working with Lesbian and Gay Male Clients: The Influence of Religious Practices and Support for Lesbian and Gay Male Human Rights." In *Journal of Homosexuality*, 2010: 57:10, 1258-1273.

Herek, Gregory M. "Assessing heterosexuals' attitudes toward lesbians and gay men: A review of empirical research with the ATLG scale." In *Lesbian and gay psychology: Theory, research, and clinical applications*, by Beverly Greene and Gregory Herek, 206-228. Thousand Oaks: Sage, 1994.

National Gay and Lesbian Task Force. "Anti-gay violence, victimization and defamation in 1989." Washington D.C., 1990. *National Gay and Lesbian Task Force Annual Report*. 2010. http://www.thetaskforce.org/about_us/annual_report (accessed June 11, 2011).

Putnam, Robert & Campbell, David. *American Grace*. New York: Simon & Schuster, 2010

Yang, Alan. "Trends: Attitudes toward homosexuality." *Public Opinion Quarterly*, 1997: 61, 477-507.

Appendix

The following data represent the first two questions in the questionnaire. These tables express the overarching attitudes towards homosexuals held by the Salt Lake City sample and the rationale behind those attitudes.

Table 1. Answer totals from all respondents to questions from the ATLG scale and Gallup polls

What are your opinions about homosexuals?		Scale*				Avg Score
		1 (%)	2 (%)	3 (%)	4 (%)	
Homosexual behavior between two men is just plain wrong.	Assailants	4.5	27.3	13.6	54.5	3.1
	Name-callers	30.2	16.7	24.0	29.2	2.5
	Non-assailants	33.4	18.8	17.7	30.0	2.4
Male homosexuality is merely a different kind of lifestyle that should <u>not</u> be condemned.	Assailants	27.3	22.7	27.3	22.7	2.4
	Name-callers	24.2	25.3	21.1	29.5	2.5
	Non-assailants	22.7	17.2	21.3	38.8	2.7
I think male homosexuals are disgusting.	Assailants	9.5	14.3	52.4	23.8	2.9
	Name-callers	39.6	20.8	16.7	22.9	2.2
	Non-assailants	51.0	17.8	15.8	15.4	1.9
Male homosexuality is a perversion.	Assailants	19.0	19.0	38.1	23.8	2.6
	Name-callers	35.1	14.9	25.5	24.5	2.3
	Non-assailants	41.4	19.7	16.9	22.1	2.1
Male homosexuality is a natural expression of sexuality in men.	Assailants	52.4	19.0	14.3	14.3	1.9
	Name-callers	37.5	21.9	26.0	14.6	2.1
	Non-assailants	33.3	18.2	26.1	22.3	2.3
Homosexual behavior between two women is just plain wrong.	Assailants	13.6	31.8	18.2	36.4	2.7
	Name-callers	34.7	14.7	23.2	27.4	2.4
	Non-assailants	36.5	20.5	14.7	28.3	2.3

Female homosexuality is merely a different kind of lifestyle that should <u>not</u> be condemned.	Assailants	27.3	22.7	27.3	22.7	2.4
	Name-callers	28.4	23.2	18.9	29.5	2.4
	Non-assailants	23.3	18.5	19.5	38.7	2.7
I think lesbians are disgusting.	Assailants	22.7	31.8	18.2	27.3	2.5
	Name-callers	43.8	20.8	15.6	19.8	2.1
	Non-assailants	51.2	19.8	13.7	15.4	1.9
Female homosexuality is a perversion.	Assailants	23.8	33.3	28.6	14.3	2.3
	Name-callers	34.0	21.3	22.3	22.3	2.3
	Non-assailants	42.8	17.5	20.5	19.2	2.1
Female homosexuality is a natural expression of sexuality in women.	Assailants	50.0	31.8	9.1	9.1	1.7
	Name-callers	31.3	20.8	27.1	20.8	2.3
	Non-assailants	32.8	18.3	25.2	23.8	2.4
Homosexuals should have equal rights to job opportunities.	Assailants	13.6	13.6	18.2	54.5	3.1
	Name-callers	5.2	5.2	18.8	70.8	3.5
	Non-assailants	3.8	1.7	13.0	81.6	3.7
Homosexual relations between consenting adults should be legal.	Assailants	36.4	9.1	36.4	18.2	2.3
	Name-callers	17.9	16.8	20.0	45.3	2.9
	Non-assailants	18.8	16.4	14.7	50.0	2.9

* 1 = strongly disagree; 2 = somewhat disagree; 3 = somewhat agree; 4 = strongly agree

ATLG = Attitudes Toward Lesbians and Gay Men Scale

Table 2. Answer totals from all respondents to Attitude Functions Inventory items

Why do you hold your opinions?		Scale*				Avg. Score
		1 (%)	2 (%)	3 (%)	4 (%)	
Because of a friend or family member who is gay or lesbian.	Assailants	63.6	18.2	9.1	9.1	1.6
	Name-callers	34.4	25.0	20.8	19.8	2.2
	Non-assailants	36.6	22.0	19.0	22.4	2.2
Because of my personal contacts with specific gay persons.	Assailants	31.8	31.8	27.3	9.1	2.1
	Name-callers	25.3	25.3	29.5	20.0	2.4
	Non-assailants	25.4	25.8	25.4	23.4	2.4
Because of knowing people who have gay family members or friends.	Assailants	31.8	36.4	22.7	9.1	2.0
	Name-callers	29.8	29.8	20.2	20.2	2.3
	Non-assailants	28.8	22.7	24.1	24.4	2.4
Because of how my friends or family feel about gay men and lesbians.	Assailants	40.9	22.7	22.7	13.6	2.0
	Name-callers	34.4	27.1	21.9	16.7	2.2
	Non-assailants	38.0	26.4	21.0	14.6	2.1
Because of how people I respect feel about gay men and lesbians.	Assailants	40.9	31.8	22.7	4.5	1.9
	Name-callers	37.5	21.9	25.0	15.6	2.1
	Non-assailants	31.5	24.7	23.4	20.3	2.3
Because I don't want to think about homosexuality or gay people.	Assailants	27.3	31.8	18.2	22.7	2.3
	Name-callers	51.0	25.0	9.4	14.6	1.8
	Non-assailants	61.7	16.6	12.2	9.5	1.6
Because of my personal feelings of discomfort or revulsion with homosexuality.	Assailants	27.3	13.6	27.3	31.8	2.6
	Name-callers	39.6	24.0	12.5	24.0	2.2
	Non-assailants	52.5	21.7	10.8	14.9	1.8

Because of my beliefs about civil liberties.	Assailants	18.2	22.7	31.8	27.3	2.6
	Name-callers	15.6	14.6	29.2	40.6	2.9
	Non-assailants	20.4	17.3	22.8	39.5	2.8
Because of my moral or religious beliefs.	Assailants	18.2	22.7	36.4	22.7	2.6
	Name-callers	28.1	9.4	20.8	41.7	2.7
	Non-assailants	25.1	14.6	13.9	46.4	2.8

*1 = not at all true of me; 2 = slightly true of me; 3 = fairly true of me; 4 = very true of me

Abstracts

ARTS

Participant Motivation in College Ballroom Dance Classes

Johnny Ahn

Utah Valley University

The rise of popularity in ballroom dance in the past 5 years has led to the proliferation of ballroom dance classes at the college level. This study interrogates why students choose to participate in these classes, at all levels. The focus of my work is Utah Valley University, which runs the second largest ballroom dance program in the nation. In a similar fashion to Finnish scholar Piepsa Niemanin, participants in ballroom classes were surveyed about why they took the class in which they were currently enrolled. The study found that rather than the anticipated socializing being the primary factor in participation, students had a greater desire to learn and improve figures and technique. I hope to expand this study to other colleges with ballroom dance programs and believe that the result will aid dance programs in better marketing and adapting their programs to attract the maximum possible participation.

ARTS

Gender Performance and Performativity in Ballroom Dance

Veronica Argyle

Utah Valley University

The purpose of this study is to explore the interplay between the performance and performativity of gender in DanceSport. Ballroom dance generally emphasizes masculine and feminine roles in pedagogy, performance, and costuming; therefore, competitors in this gendered genre have a vested interest in adhering to these roles in order to place well in

competition. This study seeks to determine whether positive reinforcement of gendered performances in DanceSport creates a performative relationship between a dancer's performance and her actual gender identity. The study will also analyze whether acknowledging the performative system within DanceSport (and society) might allow the dancer to subvert the performative system through the exaggerated performance of gender in DanceSport, thereby exposing gender as a construct and lessening its performative power. Through the works of Judith Butler and Jonathon S. Marion and personal interviews with DanceSport competitors, this study will show the complex interplay found between performance and performativity in DanceSport.

ARTS

John Wayne, Producer

Stephen Armstrong

Dixie State College

This paper will trace John Wayne's emergence as a producer of motion pictures in the 1940s and 1950s, giving particular attention to the three independent production companies he created, Patnel, Wayne-Fellows, and Batjac. The paper will consider Wayne's reasons for becoming a producer and argue that in this role he was able to develop several features that gave voice to his frequently libertarian views on topics ranging from private ownership and communism to land conservation and racism. Films cited will include *Angel and The Badman* (1947), *The High and The Mighty* (1954), and *Blood Alley* (1955).

ARTS

Lacuna

Angela Banchero-Kelleher, Amy Markgraf Jacobson, and Nichole Ortega

Utah Valley University

Lacuna, a gap. A gap exists in our relationship to the environment. We have lost our connection to the idea of ourselves as stewards of the earth. This gap in our understanding of our ritual, environmental, psychic, and mythological connection to water impedes our ability to make

informed decisions about this essential element. The Lacuna Project attempted to address this disconnect from our environment and engage members of the community in order to inspire action and partnerships and to expand a community's sense of sustainable "place." This project revealed the many nuanced approaches to issues surrounding sustainability that had ethical implications for project participants. The purpose of this presentation is to explore several facets of the process that emerged during the development of the choreographic piece, *Lacuna*, focusing on the actual process itself and the ethical issues that emerged. An exploration of these issues will inform an understanding of the role that art and the creative process play in our relationship to the environment that will help citizens validate who they are and how they relate to the life of the community and the world at large.

ARTS

International Style v. American Style: The Coaches' Perspective

Emily Darby

Utah Valley University

In the U.S. Dancesport community (the competitive ballroom dance establishment), there is an ongoing debate over which style holds more merit, the International ballroom and Latin styles or the American Smooth and Rhythm styles. The disparaging of American style, it seems, has led to less participation at the competitive level, especially in the amateur levels. On the other hand, American is more widely taught and danced for recreational purposes than is International style. In order to better understand the lack of competitive participation in American style, this study asked coaches in Utah, which has a thriving amateur competitive community, why they believe this difference in numbers of competitive couples exists in this state. A variety of reasons were posited, but one of the main opinions articulated was that there is a lack of trained American style coaches, especially in Rhythm, in Utah. Further, top officials and coaches perpetuate the idea that American style is second-best and what people do when they cannot succeed in International style.

ARTS

Dreams

Melanie Ewell Francom

Utah Valley University

To paraphrase Picasso, “dreams” are the lies that tell the truth. Sigmund Freud suggests dreams are the unfulfilled wishes, desires, and emotions of one’s mind. Using symbols in an individual’s dream, Freud’s dream theory helps to identify the meaning of the symbol and uses them to discover what an individual might be repressing from the conscious mind. Anna Sokolow, an American dancer and choreographer, created a Modern dance titled *Dreams* in 1961. Later she recognized that this particular work was more a resemblance of nightmares than of good-natured dreams as the title may suggest. A critical analysis of *Dreams* using Freud’s dream theory will reveal a greater understanding of the artist’s choreographic intent. By employing Freud’s dream theory when analyzing Sokolow’s work in *Dreams*, one can begin to uncover the symbolism behind her choreographic movement motifs. Additionally, this research explores what this work reveals about the political and social culture in America during the time that the work was created.

ARTS

The Shadow Effect

Melanie Ewell Francom

Utah Valley University

Art has been accused of imitating life, because the artist creates art through his or her own experiences. Carl Jung, the theorist and philosopher, believes art generates deep in our inner psyche. By searching within one’s self and acknowledging the shadow, one can begin to gain understanding of the collective unconsciousness that allows a place for art. The purpose of this research is to generate choreography and a work that will support discovery of the collective unconsciousness and one of Jung’s Dream Theory seven archetypes: The Shadow. The information led the choreographer to discover that to make change by facing one’s shadow takes courage, a leap of faith, so change can occur. The choreographic process can be a similar process. You conceive an

idea, explore, and develop movement to support your idea. If the choreographer takes a leap of faith, like the shadow, the work acts as a mirror, reflecting where new development can allow for something meaningful can happen.

ARTS

Reggaeton from a Feminist Perspective

Giselle Fernandez

Utah Valley University

Reggaeton sub-culture is a fertile ground for women's objectification, idolization of men's power, and control over others. Men are the singers and women are the displayers of skin. Women are subtly swayed to perform as sexual objects with the elusive hope of gaining power. From a feminist perspective, Reggaeton is a realm dominated by the male chauvinist and is the antithesis of Feminism's aims for the equality and opportunities for all; by liberating minds from stereotyping, discrimination, and sexual objectification. Reggaeton dance discriminates and stereotypes women as mere sexual objects while intending to be a sexually pleasurable performance to men; but not necessarily to the female dancer. The purpose of this research is to examine Reggaeton from a feminist perspective in order to demonstrate that it is reflective of women's objectification and the inequality between men and women in the Puerto Rican power structure currently present.

ARTS

The Disillusioned Hero: Friz Lang and Dave Bannion

Trisha Harber

Dixie State College

Politically opposed to the Nazi Regime, Fritz Lang took exile in the United States. Upon arriving in Hollywood in 1934, Lang felt that tenants of the fascist nation he'd left behind, such as censorship, were threatening freedom in his new nation. Displaced and disillusioned, Lang's films in years subsequent to his exile, such as *Fury*, *Manhunt*,

and *Ministry of Fear*, illustrate the director's distrust of and frustration toward apathetic, ineffective, and corrupt bureaucratic institutions. Dave Bannion, the protagonist of Lang's 1935 *noir* classic, *The Big Heat*, struggles with these same emotions. Ostracized, Bannion must act independently from an ineffective, corrupt bureaucracy to protect his family and freedom. Bannion's psychological state mirrors Lang's own emotional turmoil, and his vigilante response to the injustices he encounters might be an expression of Lang's own fantasies, a statement about American political apathy and his desire to strike back against the fascist intuitions of Germany.

ARTS

Night Journey: Insight to the Subconscious of Martha Graham

Lauren Harris

Utah Valley University

The creative process is one of mystery. Ideas and experiences drawn from within the innermost workings of the choreographer's subconscious are molded and shaped into works of art. The epiphanies of the creative process and the motivations that enlighten them are ambiguous and unknown. It is from the deepest reaches of one's soul that intricate and exquisite—although frequently misunderstood—expressions of humanity and culture are woven and create windows of insight into a choreographer's internal make-up. The artist turns from the realities of the world and “allows full play to his erotic and ambitious wishes” of the subconscious (Freud). It is through this “play” that the choreographer faces the unknown of the human psyche in order to discover and express the deepest levels of human emotion. A psychoanalytical perspective identifies the inspiration and creativity of the choreographer's work to be rooted within the instinctual and truest material of the subconscious and, therefore, sheds light on the creative process of choreographers in the modern era such as Martha Graham. A critical analysis of Martha Graham's *Night Journey* from a psychoanalytical perspective provides delicate sapience into the intense inner subconscious motivations that inform her choreographic work.

ARTS

Visual Scripting**Beth Miklavcic***University of Utah*

The purpose of *Visual Scripting* is to provide a complete visualization of an original script. With the advent of accessible media, especially software such as Adobe Acrobat, Adobe Photoshop, video streaming, and more, it is possible to create a multimedia script that can be provided to actors, both as drafts during the development process and completed versions for re-staging. The methods of documenting artistic creations of the present need to reflect the complicated nature of newly created original works. *Visual Scripting* documentation includes text, auditory, and visual files specifically to provide complete documentation of a new original multimedia work and to provide a foundation for interpretation by a new cast, not for the purposes of rote restaging, but for use as a foundation to move beyond the interpretations of the original cast toward finding deeper levels within the work.

ARTS

Alwin Nikolais: How His Choreography Reflects His Life Experiences**Breanna Orr***Utah Valley University*

Alwin Nikolais is recognized in Modern dance for his extensive use of props, lighting, and music. As an innovative choreographer, teacher, composer, designer, and director, Nikolais was able to create full dance theater productions that revolutionized modern dance. A critical analysis of Alwin Nikolais's choreography as a whole, as well as an in-depth look at *Tensile Involvement* employing a psychoanalytical frame of analysis, will provide insight into some of the motivations that directed his artistic vision. This paper will first define Nikolais's artistic philosophy, and second, explore his specific life experiences involving theater, puppeteering, piano, World War II, and his own homosexuality that informed his choreographic motives. Freud's theories of behavior and the unconscious will be used to provide a deeper understanding of the wonderful works of Alwin Nikolais. This synthesis will create new

knowledge about how works such as *Tensile Involvement* display his philosophy about decentralization, dehumanization, and desexualization of the body and directly reflect his life experiences.

ARTS

Rudolf von Laban and the Nazis

Aleisha Paspuel

Utah Valley University

Looking at the relationship between the emerging Modern Dance in Germany and the Nazis, it becomes clear that the mass movement style in German Modern Dance correlates to the vigorous promotion of a unified national community by the Nazis. Yet there has been the idea that German Modern Dance developed in spite of the Nazi movement and that many German dance artists had to work in secret. This speculative notion is incorrect. Not only did many German dance artists openly work in Germany, many even supported the Nazis through their work. Rudolf von Laban was a prime example; his work reflected Nazi ideology, and under them he successfully developed his work in German Modern Dance. Using an historical frame of analysis, this research will discuss the connection between Rudolf von Laban's movement choirs and Nazi ideology, demonstrating the close ties between early German Modern Dance and the Nazi Party.

ARTS

Going Through Life Sideways: Fritz Lang's Violence with his Female Characters

Elaine Sharp

Dixie State College

This paper will analyze the thematic function of violence as it is employed against female characters in Lang's cinema. The paper will consider the personal relationships the director had with the women in his life, particularly his three wives, as well as his mother and grandmother. It will then suggest that Lang's frustrations with these individuals found expression in films like *Metropolis* (1927), *Clash by Night* (1952), and *The Big Heat* (1954). Moreover, it contends that

Lang's violent treatment of female characters demonstrates a distinct repulsion toward women, despite claims he made to the contrary.

ARTS

Backwards and in High Heels: The Myth of Female Passivity in Ballroom Dance

Roger Wiblin

Utah Valley University

Some non-ballroom dancers have had conjured up for them the idea that in ballroom dance men have all of the power and women willingly and passively submit to whatever he desires. The famous quote about Ginger Rogers referenced in the title, together with popular television shows and movies about ballroom dance and some recent ethnographic ballroom scholarship, have helped to perpetuate this stereotype. This notion is a false one. Through interviews with Utah coaches and competitors, this study demonstrates that the female partner has significant agency to choose whether and how to respond to the lead given by her partner. Further, while the man always leads, he is often not the source of power as the partnership moves across the floor. It will be demonstrated that the lady plays a much more significant and active role than she is often given credit for.

BIOLOGICAL SCIENCES

Comparison of *Idiomarina* Bacteriophage Isolated from the Great Salt Lake, Utah

Carlie Benson, Craig Oberg, Matthew Domek, and Michele Culumber

Weber State University

Predation by bacteriophages (phage) may play a significant role in controlling bacterial populations in hypersaline environments. Prior characterization of halophilic bacteria from the South Arm of the Great Salt Lake (GSL) revealed a number of novel *Idiomarina*-like isolates (ILI). The three ILI hosts (SA03, SA06, and SA11) were nearly 99.5% identical over the 16S rRNA gene. Using the ILI as hosts, phages were

obtained from GSL water. Host range analysis showed that phage CB04 only infected SA11 while phage CB13 only infected SA03. One purpose of this research was to demonstrate that different phage can selectively infect nearly identical ILI strains. Nucleic acid was extracted from CB04 and CB13 and enzymatic digests revealed both have DNA genomes between 50 and 55 kb. However, restriction enzyme digest patterns of the phage showed no similarity. These observations could indicate the need for increased specific phage predation during times of seasonal high cell density, competition between closely related species, or acquisition of phage resistance by some ILIs.

BIOLOGICAL SCIENCES

A Unified Study of Ethics on Conservation of Biodiversity

Ruhul Kuddus

Utah Valley University

All members of a species living in an environment constitute a biological population, and all populations living in an environment constitute a biological community. A community is the smallest unit that can thrive in a natural environment. Conserving life of an individual human being is a high-priority ethical burden on individuals as well as the society. An individual of a species has a natural lifespan of a few minutes to a few thousand years. In contrast, the estimated natural lifespan of a typical species is about five million years. Premature extinction of a non-human species can be detrimental to the well-being of a biological community and the human species. Death of a species because of human activities presents a severe ethical dilemma. The present study applied secular and religious (mostly Islamic) ethical principles on conservation of life and came to a logical conclusion that the ethical burden of conserving life of a species of a biological community is as important as conserving life of an individual human being. Ethical issues of common conservation practices, such as creating strictly protected habitats and species, are addressed. The study indicated that modern technologies such as tagging, cloning, and genetic engineering can be judiciously applied in conserving non-human species without violating secular or religious ethical norms.

BIOLOGICAL SCIENCES

***Batrachochytrium dendrobatidis* fungus update for populations of *Hyla arenicolor* in Zion National Park**

Britnee Moore and Curt Walker

Dixie State College of Utah

Batrachochytrium dendrobatidis is a form of chytrid fungus that results in a condition known as chytridiomycosis. It is a contributing factor in decreasing global amphibian populations. To date, chytridiomycosis has been found on all continents and in several areas within the United States. Data from the summer of 2009 indicated that chytridiomycosis was not found in the canyon tree frog, *Hyla arenicolor* within Zion National Park in southwestern Utah. To determine if these populations were still free of chytridiomycosis one year later, samples were collected in the summer of 2010 and tested for the presence of *B. dendrobatidis*. The results help us to establish a timeline for the spread of *B. dendrobatidis* into the amphibian populations of Zion National Park.

BIOLOGICAL SCIENCES

Isolation and Characterization of Cellulytic Microorganisms from the Great Salt Lake, Utah

Elizabeth Mora, Brian Bill, Craig Oberg, and Michele Culumber

Weber State University

The degradation of cellulose to sugars for ethanol fermentation by unique cellulases may play a role in the development of alternative fuel sources. The Great Salt Lake (GSL) provides a opportunity to isolate microorganisms and enzymes that may have a unique ability to utilize cellulose as their carbon source. GSL samples of water and sediment were inoculated into a halophilic minimal medium enriched with powdered cellulose, grass clippings, or newspaper. After 3 to 4 weeks, microorganisms were isolated from the enrichments on halophile agar plates containing cellulose. After isolation, the 16SrRNA gene was sequenced. On all of the substrates, Halomonas-like organisms were isolated. An Idiomarina-like organism was isolated from the grass clipping enrichment. The cellulytic activity of individual isolates was quantified using a CMC-iodine cellulase assay. Several isolates exhibit

moderate exo-cellulase activity warranting specific enzyme isolation and characterization.

BIOLOGICAL SCIENCES

The Effect of Rock Pool Properties on the Diversity and Abundance of Microflora and Fauna in SW Utah Slot Canyon Pools

Chad Roberts

Dixie State College

Chemical and physical properties of rock pools in red sandstone slot canyons of Zion National Park in southwestern Utah were measured to determine if they affected the microscopic biological communities of those pools. In other parts of Utah, ammonium concentration shows a strong positive correlation with both the diversity and abundance of species in slot canyon pools. Ammonium concentration was measured in addition to water temperature, water movement, pH, and light availability in the pools of 9 slot canyons over a 7-month period. Only a few of these variables correlated with the composition of microscopic biological communities. The presence and frequency of two different groups of organisms were positively correlated with ammonium concentration and light availability, while a third was positively correlated with water temperature and humidity. The results suggest that the effect of ammonium concentration in determining species diversity found in rock pools at other Utah sites appears to be important for some species but does not universally affect all species, while other factors can also affect species composition of the microorganismal communities in slot canyons and should be included in future research.

BIOLOGICAL SCIENCES

Novel *Marinobacter*-like Organism and a Related Phage Isolated from the Great Salt Lake

T. B. Simon, C. J. Oberg, M. D. Culumber, and M. J. Domek

Weber State University

The diversity of microorganisms of the Great Salt Lake (GSL) remains largely unexplored, with the most easily cultivated bacteria belonging to the γ -Proteobacteria genera *Salinivibrio*, *Idiomarina*, and *Halomonas*. We recently isolated a *Marinobacter*-like organism along with a halophilic phage that infects it. GSL water samples were diluted, plated on an oligotrophic halophilic medium, and incubated at 30°C. A small, slow-growing, colony was selected, cultivated in broth, and identified by its 16S rRNA gene sequence. Growth was optimum at 4% NaCl (range 4-12% NaCl). Phage were enriched for by mixing filtered GSL water with the host and then isolated using a soft-agar plaque assay. The phage formed 4-5-mm diameter plaques in 0.8% soft agar. Although *Marinobacter*-like organisms have not been previously identified in the GSL, the presence of a specific bacteriophage for this isolate suggests *Marinobacter* is relatively abundant in GSL water and may be vital to the GSL ecosystem. Other members of the *Marinobacter* are known to metabolize hydrocarbons, a process that may be important in the GSL.

BIOLOGICAL SCIENCES

Antibacterial Properties of *Origanum vulgare* Essential Oil on *Staphylococcus aureus*

Marcus Stucki, Dennis Farnsworth, Zachary Hazenwinkel, and Don Warner

Dixie State College

The biological effects of food spices have been studied for many years. Of particular interest is the inhibitory effect of food spices on bacterial pathogens. The inhibitory effect of essential oil (EO) from *Origanum vulgare* (oregano) on *Staphylococcus aureus* was tested using previous methodology to determine the validity of this test procedure. The minimum inhibitory concentration of oregano EO was determined by (1) measurement of optical density of a 1/2 serial dilution of oregano EO

and (2) measurement of color change of p-iodonitro tetrazolium dye of a 1/2 serial dilution of oregano EO. Repeated measurements using these methods were made, and the results statistically analyzed and statistically correlated. Both methods showed positive inhibitory effects of oregano EO on *S. aureus* at approximately the same dilution.

BIOLOGICAL SCIENCES

Susceptibility Study of *Staphylococcus aureus* Carotenoid-Deficient Mutants to Oxacillin

Uriwan Vijaranakul

Pibulsongkram Rajabhat University

R. K. Jayaswal and Brian J. Wilkinson

Illinois State University

The carotenoid pigments of *S. aureus* play a significant role in resistance to host defense systems and provide integrity to its cell membrane. However, the role of carotenoids in susceptibility of *S. aureus* to antimicrobials has not been studied extensively. This study has been conducted to examine the impact of staphyloxanthin on the in vitro antimicrobial susceptibility of *S. aureus*. The carotenoid deficient Δ crtM mutant, its parent strain and crtM complemented strain were analyzed for their susceptibility to oxacillin by using gradient plate technique and MIC determination by microdilution broth methods. Both gradient plates and MIC showed that Romero Δ crtM mutant has decreased susceptibility to oxacillin.

BIOLOGICAL SCIENCES

Inhibition of *Clostridium difficile* by Lactic Acid Bacteria

Tarris Webber, Jed Lowe, Rachel Lowe, and Craig Oberg

Weber State University

Clostridium difficile is a spore-forming, anaerobic bacillus that accounts for 15-25% of all episodes of antibiotic-associated diarrhea. This study was done to determine if lactic acid bacteria (LAB) used in

dairy products have an inhibitory effect against *C. difficile*. With the difficulty of propagating *C. difficile* on solid media, 3 media types (sheep blood agar, sheep blood agar with oxyrase, and MRS agar) were tested. All LABs grew on these media with *C. difficile* showing the most rapid growth on oxyrase-sheep blood agar. Sixteen LAB cultures were challenged with *C. difficile* using the agar flip technique. Of the 16 LAB strains used, seven inhibited *C. difficile*. Most LABs belonged to the genus *Lactobacillus*. These results show that species of LAB can inhibit *C. difficile*. Since there was no direct contact between the two bacteria, inhibition may be due to secreted LAB metabolites including organic acids or bacteriocins. Results show that LAB strains, delivered in fermented dairy products, have potential as a treatment for *C. difficile* infections.

BUSINESS

Right to Work Policies in Utah

Troy Callahan

Utah Valley University

This paper will explore several issues pertaining to the “right to work” law in Utah. These issues include the definition of the law, employees’ understanding of the law, and whether or not there exists a fundamental misunderstanding of the law that contributes to a culture of exploitation of employees by their employers. For example, I asked my 14-year-old son to tell me what he thought “right to work” meant. He thought for a minute and answered that it meant that he had the right to work, if he wanted to. I decided to randomly ask the same question of different people, and most had a different answer. I asked a woman in a convenience store what she thought “right to work” meant, and she replied that it meant her boss could “fire her ass” whenever he wanted to. Most people I questioned had the same understanding that the convenience store clerk had. This paper will accurately define the law and dispel the notion that “right to work” means “right to abuse.”

BUSINESS

A Note on the Usefulness the Minimum-Variance Portfolio in Practice

Leo Chan

Utah Valley University

In this study, we investigate the usefulness of the Markowitz Portfolio Theory by using data from the U.S. stock markets from 1987 to 2005. We found that it is impossible to construct an efficient frontier by static analysis of the stock market data. While the minimum variance portfolio does produce better risk-adjusted return for the investor during the 1996 to 2005 period, the result from the 1987 to 1997 period shows otherwise. Therefore, we cannot conclude that the Markowitz Portfolio Theory is of any use to an average investor. Even if institution investors have the financial resources to carry out unlimited portfolio rebalancing, the time-varying nature of correlation coefficient makes the practice impossible when transaction cost is taken into account.

BUSINESS

Partially Adaptive and Semiparametric Estimation of Regression Models for Grouped Data

Jason B. Cook and James B. McDonald

Brigham Young University

Several valuable data sources, including the Census and the National Longitudinal Survey of Youth, include data based on interval responses. Many empirical studies attempt estimation by assuming the data correspond to the interval midpoint and run ordinary least squares regressions or performing maximum likelihood estimation (MLE). Stata provides a MLE approach, based on the assumption of normality, which allows for intra-group variation. In the presence of misspecification, these estimates are inconsistent. Other solutions include utilizing nonparametric estimation techniques. In this paper, we take an intermediate position, using partially adaptive estimation, which builds on a MLE framework using flexible parametric probability density functions. Semiparametric procedures are also considered. Applications of these methods are used to estimate determinants associated with household income based on U.S. Census data. Monte Carlo simulations are

performed to evaluate the relative efficiency of the different methods of estimation.

BUSINESS

Organizational Politics in a Utah Context

Trista Doxey

Utah Valley University

Office politics is often experienced but is rarely mentioned. Political behaviors are experienced within every organization. Determining the effects and causes of organizational politics can help a company use the behaviors to increase productivity, creativity, and positivity. Many companies could benefit from having company-wide policies concerning these types of behaviors. Therefore, by prohibiting detrimental behaviors and promoting positive behaviors, a company can increase productivity, creativity, and positivity. Additionally, political behaviors differ between leaders and subordinates within an organization. Positive politics are used to benefit the organization while negative politics are often used for self-promotion. Positive political behaviors include network building and constructive discussion. Negative behaviors include deception, coalition building, and favoritism. This paper examines a local (Utah) organization to analyze how these behaviors are used and how they can be used to promote positive behaviors.

BUSINESS

Sources of Entrepreneurial Passion

Michael Glauser

Westminster College

Melissa Cardon

Pace University

Many popular and academic writings have argued somewhat convincingly that passion is an important aspect of entrepreneurial success. Passion has been said to increase an entrepreneur's commitment to the venture, persistence in pursuing venture-related goals, and ability to stay full engaged in the venture over time. However, we have very lim-

ited empirical knowledge of the sources of such passion. In this study, we analyzed the oral histories of 80 successful entrepreneurs from a variety of businesses and industries to explore the sources and sustenance of their passion. Our findings suggest there are 6 major sources of entrepreneurial passion: passion for building/developing the venture, passion for people, passion for the product or service, passion for inventing, passion for success, and passion for a social cause. These results show that entrepreneurial passion is fairly specific rather than a general enthusiasm for the venture, as has been previously suggested. We present implications for future research on the relationship between the various sources of passion and entrepreneurial commitment, energy, tenacity, and the overall success of new ventures.

BUSINESS

What Successful Entrepreneurs Really Do

Michael Glauser

Westminster College

Entrepreneurial activity is increasing throughout the United States and the world. Despite this growing enthusiasm for entrepreneurship, the failure rate for new ventures remains high. Although we are learning more about the enigmatic process of launching new enterprise, a deeper understanding of the complex human drama from the perspective of those who actually do it well will shed greater light on factors associated with start-up success. The study analyzed the oral histories of several hundred entrepreneurs from a variety of businesses in a variety of industries. The stories were collected from 1994 to 2010. Through a detailed content analysis, factors were identified that appeared to have a significant impact on the success of each company. Categories were created to contain the factors and then collapsed through an iterative process into key principles common to a majority of the ventures. These practices constitute an entrepreneurial leadership model for the 21st century. The model has obvious implications for aspiring entrepreneurs, investors, corporate executives, government leaders, and researchers in the field of entrepreneurship.

BUSINESS**Advertising to Children: A Review to Determine Ethics and Social Responsibility****Benjamin Hart***Utah Valley University*

In the context of television advertising, parents and businesses have conflicting opinions about the effects of advertising on children and whether children's advertising is ethical. The assignment of responsibility for the protection of children remains subject to debate. This paper reviews the issues surrounding the conflicting opinions and purposes to address any revealed ethical concerns that will help in determining whether children should be protected from television advertising and which part of society is best suited for that responsibility. Literature is reviewed from research articles in advertising, marketing, health, and developmental psychology. The review concludes that parents are needed to educate children about products that do not promote the public good, and advertisers are qualified to identify needs for reform and have the ability to make changes. Future research will help advertisers identify factors that contribute to the unfavorable perception of advertising to children and give them direction for reform.

BUSINESS**Diversification and Comparative Advantage: 160 Years of Change in Utah's Economy****L. Dwight Israelsen***Utah State University*

Economic theory suggests that gains from trade are maximized when an individual, region, or country specializes in and exports products in which the entity enjoys a comparative advantage. Without government interference in the form of trade barriers, self-interest and competitive markets will ensure that comparative advantage is, indeed, the primary basis of trade patterns and, hence, economic development patterns. Utah's history provides an interesting example of a pattern of economic development in the 19th century that deviated from comparative advantage in order to achieve self-sufficiency, resulting in an economic structure that more closely resembled that of the entire U.S. economy than

of its neighboring states. This diversified economy helped protect Utah from the economic instability associated with swings in world metal prices that destabilized the economies of other Western states. While Utah's economy came to more closely resemble that of surrounding states between the 1880s and 1960s, and was correspondingly less stable than the overall U.S. economy, changes in Utah's comparative advantage in the last half century have led to a more stable, diversified economic structure that once again closely resembles that of the U.S. as a whole.

BUSINESS

An Analytical View of China as a Software Outsourcing Outlet

Taowen Le

Weber State University

Wayne Huang

Ohio University

Jin Zhang

University of Wisconsin Milwaukee

This article analyzes the potential of China as a software outsourcing outlet. It reviews the progress China's software industry made in recent years, discusses hindering factors that once constrained China's potential as a software outsourcing service provider, and identifies various enabling factors that contributed to China's growing potential as a software outsourcing service provider. It analyzes the potential impact China's growing capabilities as a software outsourcing outlet may have internationally and identifies what China would need to do to become a greater contributor to the world's software outsourcing service industry. Although the study focuses on China, its analysis and suggestions can be valuable to other developing countries as well.

BUSINESS**Partially Adaptive Estimation of the Censored Regression Model****Randall A. Lewis***Massachusetts Institute of Technology***James B. McDonald***Brigham Young University*

Data censoring causes ordinary least squares estimates of linear models to be biased and inconsistent. Tobit, nonparametric, and partially adaptive estimators have been considered as possible solutions. We propose several new partially adaptive estimators that cover a wide range of distributional characteristics. A simulation study is used to investigate the estimators' relative efficiency in these settings. The partially adaptive censored regression estimators have little efficiency loss for censored normal errors and outperform Tobit and nonparametric estimators for non-normal distributions. An empirical example of out-of-pocket expenditures for a health insurance plan supports these results.

BUSINESS**The Impact of Workplace Wellness Programs on Decreasing Employee Obesity and Increasing Overall Health****Skyler C. Macdonald and Jonathan H. Westover***Utah Valley University*

Obesity is a deadly disease that is spreading across America and creating an epidemic that needs to be stopped. The risks are very serious and have a great impact on the quality of the lives of those with obesity. The obese individual is not the only one who suffers the consequences of their obesity. The whole of society is being affected by this epidemic. The costs of obesity affect not only the individual but the company that employs him or her. A solution to obesity is through creating a corporate culture that is physically active where the company provides employee benefits that not only increase the well-being of the employees but also improve the quality of their lives. Corporate wellness programs benefit the whole of a business both financially and

through the health of the employees. It is important to a company to improve and sustain the health of its employees. The implementation of wellness programs can influence the employees in living healthier lives. There needs to be a change. Employers are in the best position to make the largest impact on the greatest amount of people.

BUSINESS

Moments to Remember: Some Popular Models for the Distribution of Income

James B. McDonald, Jeffrey T. Sorensen

Brigham Young University

Patrick A. Turley

Harvard University

This paper explores the ability of some popular income distributions to model observed skewness and kurtosis. We present the generalized beta type 1 (GB1) and type 2 (GB2) distributions' skewness-kurtosis spaces and clarify and expand on previously known results on other distributions' skewness-kurtosis spaces. Data from the Luxembourg Income Study are used to estimate sample moments and explore the ability of the generalized gamma, Dagum, Singh-Maddala, beta of the first kind, beta of the second kind, GB1, and GB2 distributions to accommodate the skewness and kurtosis values. The GB2 has the flexibility to accurately describe the observed skewness and kurtosis.

BUSINESS

The Challenge: Experiential Education in Theory and Practice

Peter Robinson, Laurent Josien, and Rachel McGovern

Utah Valley University

Because of the complex nature of entrepreneurship and the wide range of knowledge, behaviors, and motivations involved in the entrepreneurial processes, experiential education has emerged as the foundation pedagogy for developing entrepreneurs. It has become critical to pro-

vide theory-backed models and principles as a foundation for the pedagogy. This paper explores the foundations and structure of attitude theories and goes on to provide a model for the application of attitude theory in experiential education within the context of entrepreneurship education. The application of theory is illustrated through an in-class entrepreneurship simulation referred to as “The Challenge.”

EDUCATION

Practical Theory: Making Research Writing Theory Useful for Students

Angie Carter

Utah Valley University

Based on Doug Brent’s (*Reading as Rhetorical Invention: Knowledge, Persuasion, and the Teaching of Research-Based Writing*) and Margaret Kantz’s (“Helping Students Use Textual Sources Persuasively.” *College English* 52.1) claims that rhetorical reading forms the basis for argumentative writing, this paper argues that, for students to most effectively complete their research papers, the practical aspects of doing research must be connected with the theoretical underpinnings of why we research. For teachers, I offer specific ways to connect theory with practice. For example, although I discuss timing, scaffolding, and drafting, I stress the importance of practical classroom activities that make the “nuts and bolts” of research writing actually useful for students. Along the way, I offer specific suggestions on how the theoretical can be made more visible to our students through the practice we provide. In my experience, both as a researcher and a teacher, I find that the more practical the activities I provide in class, the more successful my students are with their papers on their own.

EDUCATION

Contracting for Better Grades

Aaron K. Combs

Dixie State College

Student retention is an important topic for colleges and universities. One variable that has been shown to affect college retention is grades.

Implementing behavioral modification techniques to train study skills has been effective in increasing student success, but these techniques are currently not used on a regular basis. If behavioral modification techniques such as behavioral contracts are implemented in introductory classes, students may develop more effective study habits, and their grades may improve. In this study, I examine the techniques that teachers of Psychology 1010 courses at Dixie State College use to teach students successful study skills. I present research illustrating how much the average student studies in those courses, and then I compare the results to students in comparable courses who receive behavioral contract implementation. This presentation also models effective implementation of behavioral contracts that promote early study skills.

EDUCATION

Mobile Technology

Talitha Hudgins

Utah Valley University

The last decade has witnessed an incredible advance in broadband and wireless technologies. Depending on these developments, from business to education, various sectors try to use Internet and mobile devices intensively to increase their effectiveness and efficiency. Some educators, even though they are accused of being resistant to innovations, are eager to adopt these new technologies in educational settings, where mobile technologies seem to be the unique media to satisfy the slogan 'whenever and wherever.' Mobile devices are the only technologies that can be with us everywhere and at all times. In comparison, traditional technologies such as radio, television and computers have a few limitations. Radio, television, or computer cannot be carried around all the time. Furthermore, radio and television allow one-way communication without providing any interaction opportunity between a teacher and a learner. In contrast, mobile technologies can deliver education when and where it is needed. Mobile technologies can also contextualize the learning environment with its interactive multimedia presentational capabilities such as video, audio, graphics, integrated media, and two-way communication. Some disadvantages of mobile learning devices are small screen sizes and keypads. In contrast, one of the advantages of mobile technologies is that they can be complementary to the existing educational settings and materials. Mobile devices offer students the opportunity to use multimedia learning content. However,

scrolling on a small screen, where information is text-intensive, is an important factor in reducing the performance, satisfaction, and effectiveness of mobile devices for learning. In addition to auditory materials, graphics, animations, and videos can also be used to develop learning content for mobile devices. Educators using mobile devices for instruction and student learning is implementing today's digital tools for a digital generation. Using these tools encourage collaboration, sharing, teamwork, personalize learning experiences, differentiated learning, and an anywhere, anytime learning experience adapting to the needs of each learner. Using various mobile devices offers a more affordable tool than a complete laptop or desktop computing system plus the portability aspect of mobile devices. The educational implication of using such devices as the iTouch, iPad, and/or Smartphone's can deepen students' content knowledge by exposing them to high-quality visual and aural representations of places, people, events, and concepts they may not otherwise experience. Additionally, using such devices in the educational content can sharpen students' critical thinking and information literacy.

EDUCATION

Developing Instruction for the Diverse Classroom: Collegial Collaboration in Action

Vessela Ilieva

Utah Valley University

How do teachers acquire the knowledge and skills needed to adequately respond to the significant growth of ethnically, linguistically, and culturally diverse school-age children in Utah and across the United States? This research explored one model of engaging teachers of mathematics and English as a second language (ESL) in a collaborative planning of lessons that supported the academic learning of English language learners in the mainstream classroom. The teachers worked together to identify appropriate teaching strategies and techniques for classrooms where students may lack sufficient linguistic or cultural background needed to make sense of academic tasks and mathematical problems. With the help of their ESL colleague, the mathematics teachers discovered how these missing connections might prevent students from grasping important mathematical knowledge. Together, they worked on expanding their skills in engaging *all* of their students in

meaningful learning that considered their different backgrounds as part of the teaching process.

EDUCATION

Visions and Blueprints for Tomorrow's Schools: Why We Need Them Now

Jim McCoy and Prent Klag

Southern Utah University

School restructuring and reform is foremost in the minds of most educators, parents, legislators, and citizens in America. In order to be successful and make significant changes in schools, it is essential that educators, and all school stakeholders, think creatively and outside the normal realm of conditions and circumstances that exist in our schools. To make schools 'extraordinary,' concerted attention must be made to a variety of factors that mold the school environment and impact the teaching and learning that takes place. Through a synthesis of research on school reform and restructuring, and with an innovative twist on the factors that can make schools great places, this presentation will explore how a Personalized Learning Center Model would incorporate 7 key components that could make a substantial difference in schools of the future.

EDUCATION

Teacher Training at Snow College, 1888-1936

David Rosier

Snow College

With the passing of the Free Public School Act in 1890, Mormon leaders in Utah feared they were being pushed out of influence, believing control over education was essentially control over the territory. Their immediate reaction was establishing teacher training academies where-in young people of their own faith could be trained, thus retaining influence while at the same time obeying the law. Two-year-old Snow Academy very quickly adapted, beginning 46 years as an effective Normal School. Snow did this with flair, embracing many new pedagogical theories such as kindergarten and nature studies. From the be-

ginning Snow had a teachers' training lab, using children from the local public school. The result was the training of hundreds of young teachers. The program continued with notable success, firmly but indirectly maintaining Mormon leadership in the state. Snow discontinued its Normal program when requirements for teacher certificates were changed.

EDUCATION

The Use of Instructional Time for Teaching Science at the Elementary Level

Mary Sowder

Utah Valley University

Elementary teachers often receive inconsistent messages from national and local officials about the value of science instruction relative to the value of teaching subjects emphasized on standardized tests. This study looks at how these messages affect the practices of elementary science teaching as sampled in the classrooms of 179 K–6 teachers across 10 school districts in the Intermountain West. Results of this investigation confirm earlier studies about the inconsistent amounts of time and resources devoted to science instruction in elementary classrooms and call into question elementary teachers' notions about the effects of devoting class time to science teaching on subjects included in measures of annual yearly progress.

EDUCATION

Assessing Effective Instructional Design in the Developmental Mathematics Classroom

Jacque P. Westover

Utah Valley University

Increasing student success in low-level college math courses is challenging. Using an instructional design that engages students and encourages collaboration is the key to creating a problem-solving atmosphere. My efforts to have a classroom that helps students to grow mathematically as individuals and as a group were focused on one de-

velopmental math content unit covering decimals, ratios, proportions, and percents. I developed student-centered learning goals, created engaging and collaborative-learning lesson plans, and used a variety of formative and summative assessment methods to enhance and evaluate the learning of my students. Pre-test and post-test data analysis shows great improvement in the students' understanding of these mathematical concepts, as well as increased confidence in their overall mathematical problem-solving abilities.

ENGINEERING

Mars to Go Lite: A Minimal Manned Mars Mission

Donovan Chipman and David Allred

Brigham Young University

A concept to deliver one astronaut to Mars for 3-5 days and return him/her to Earth using a minimum amount of new technology is considered. This does not require new Mars entry, descent, and landing technology beyond that already qualified by NASA. Two Falcon-IX Heavy launch vehicles are required to send the following payloads to Mars: 1) a deep space-capable habitat and/or capsule with 1 astronaut; 2) two VASIMR electric propulsion tugs with 150 kW solar panel arrays; and 3) two Mars Science Laboratory (MSL) aeroshells carrying a) a hydrogen/oxygen-fueled ascent vehicle with small enclosure for one astronaut in a spacesuit, and b) a skycrane lander with a lightweight, collapsible, manned surface rover. Centaur upper stages serve as the earth departure stages for all payloads. We consider fitting landed vehicles in the limited volume of an MSL aeroshell/skycrane system and include simple interplanetary trajectory analysis for low-thrust electric propulsion.

ENGINEERING

Analytical and Experimental Evaluation of Aerodynamic Thrust Vectoring on an Aerospike Nozzle

Shannon Eilers and Matthew Wilson

Utah State University

Results from numerical and cold-flow experimental investigations of aerodynamic thrust vectoring on a small-scale aerospike thruster are presented. Thrust vectoring was created by the injection of a secondary fluid into the primary flow field normal to the nozzle axis. The experimental aerospike nozzle was truncated at 57% of its full theoretical length. Data derived from cold-flow thrust vectoring tests with carbon dioxide as the working fluid are presented. Injection points near the end of the truncated spike produced the highest force amplification factors. Explanations are given for this phenomenon. For secondary injection near the end of the aerospike, side force amplification factors up to approximately 1.4 and side force specific impulses up to approximately 55 s with main flow specific impulses clustering around 38 s were demonstrated. These forces crisply reproduce input pulses with a high degree of fidelity.

ENGINEERING

Attitude Control Using Aerodynamic Vectoring on an Aerospike Nozzle

Nathan Erni and Crystal Frazier

Utah State University

Results from a scaled small satellite hardware in the loop demonstration using an aerospike nozzle with secondary injection ports are presented. The given demonstration validates the proposed CubeSat-Scaled Propulsion System (CCSPS) for precisely positioning CubeSats to accomplish meaningful orbital maneuvers. The small form factor of the aerospike nozzle allows tactical objectives to be accomplished without the need of bulky gimbaling systems. Integration on small units such as a CubeSat become feasible due to the compact package of the designed aerospike. A one-axis demonstration provides positioning data for a carbon dioxide aerospike nozzle with an annular exhaust of

0.1 kg/s and secondary injection exhaust of 2% of the main exhaust. The aerospike is truncated to 40% of the original spike and two secondary injection sites are located near the aft end of the truncated region. Positioning tests are compared with and without annular flow. An amplification factor of 40% is realized when annular flow was present.

ENGINEERING

Mechanical Property Effects on Anisotropic Thermal Conductivity of Planetary Soils

Daniel Garrett

Utah State University

Determination of planetary heat flow values depends on the measurement of thermal conductivity of a planetary surface. The needle-like probes typically used for this measurement measure thermal conductivity in the radial direction. Thermal conductivity of planetary soils has been assumed to be isotropic, meaning the measured value of radial/horizontal thermal conductivity is the same as the thermal conductivity in the desired vertical direction. Thermal conductivity is related to the contact area between soil particles. This contact area is a function of pressure within planetary soils. The three-dimensional pressure distribution within a soil results in differing directional thermal conductivities. This research intends to show the effect of varying mechanical properties on the anisotropic compressive pressure dependent thermal conductivity of soils. This is achieved through experiments and finite element modeling of granular beds consisting of spherical particles, which are then extended to planetary soils.

ENGINEERING

Enhancing Multispectral Imagery of Ancient Documents

Trace A. Griffiths

Utah State University

Multispectral imaging (MSI) provides a wealth of imagery data that, together with modern signal processing techniques, facilitates the en-

hancement of document images. In this paper, image fusion, matched filters, bleed-through removal, and shadow removal are reviewed and applied to ancient documents. These four areas of focus provide useful tools for papyrologists studying the digital imagery of documents. The results presented form a strong case for the utility of MSI data over the use of a single image captured at any given wavelength of light.

ENGINEERING

SAM Instrument Control and Tracking

Spencer Jackson

Utah State University

SAM (Sun and Aureole Measurements) instruments were designed by Visidyne to measure the water vapor column and other atmospheric information by computations on measured data. Control software for the SAM instrument is typical of many hardware control programs, designed in an object-oriented approach, involving interfacing of devices and providing a user interface. SAM Instrument Control is unique in that it runs several control loops in order to achieve careful solar tracking and to record useful data. Employed strategies include the Hough transform and least squares estimation.

ENGINEERING

Document Flash Thermography

Cory A Larsen

Utah State University

Flash thermography, a subset of pulsed thermography or pulsed video thermography, is a technique commonly used for flaw detection in a variety of materials, including concrete, high-density polyethylene, CFRP/GFRP aerospace composites, wood and wood-based materials, and adhesive bond evaluation. This technology previously had not been applied to documents or archaeological artifacts to reveal covered writings; for instance those found in the Codex Selden or Egyptian cartonnage. Developed is the theory and application of flash thermography to documents, and a foundation is laid for the application of this technology to archaeological artifacts in general. Development of flash ther-

mography for this application provides the capability to non-destructively reveal covered writings—advancing the knowledge about ancient cultures without damaging irreplaceable artifacts.

ENGINEERING

The Effect of Surface Geometry and Heat Flux on Thin Wire Nucleate Pool Boiling of Subcooled Water in Microgravity

Troy Munro

Utah State University

In the summer of 2010, undergraduates from the USU Get Away Special team flew a nucleate pool boiling experiment on NASA's Weightless Wonder to study nucleate boiling heat transfer in microgravity. The motivation of this research is to understand the effects of surface geometry and heat flux applied to a thin wire heater. This will further the understanding of the fundamental behaviors of boiling onset, steady-state heat transfer, and bubble dynamics with respect to nucleate boiling with the goal of creating efficient thermal management systems for space applications. Using 3 thin platinum wire geometries and 5 different power levels, subcooled water was boiled over a period of approximately 30 seconds over 15 parabolic arcs to simulate microgravity. To represent the trends in bubble behavior across hundreds of frames of video in a single graph, a new method, named relative bubble area analysis, is introduced and used to analyze the results. It was determined that the efficiency of steady-state heat transfer via nucleate boiling in microgravity is comparable to, and in some cases more efficient than, steady-state heat transfer in terrestrial applications. The three-wire geometry reduced the heat flux necessary to initiate boiling. Bubble dynamics show a transition from isolated bubbles to jets of small bubbles as heat flux increases. This can be confirmed visually and with relative bubble area analysis. This implies that sustained convective heat transfer with subcooled water is possible in microgravity. A three-wire surface geometry was shown to initiate boiling at lower heat fluxes, which would provide minimal super heating of the surface, from lack of convection, before boiling heat transfer could begin.

ENGINEERING

SABER OH Mesospheric Airglow Emissions

Bryant Svedin

Utah State University

Mesospheric hydroxyl-airglow volume emission rates were derived from the NASA LaRC/USU SABER instrument aboard the TIMED satellite. The SABER 1.6- μm and 2.0- μm radiometric channels provide measurements of the OH (5,3;4,2) and (9,7;8,6) emission bands, respectively. From these data, the peak values of the volume emission rate and the altitude were ascertained for the OH layer during each valid scan. Global values were interpolated and three-dimensional global plots generated using the altitude of the OH (5,3;4,2) peak as the z-axis. The value of the OH (5,3;4,2) volume emission rate at the emission peak is displayed on the three-dimensional surface using a color scale. The ratio of the OH (5,3;4,2) band to the (9,7;8,6) band at the OH (5,3;4,2) altitude of peak emission is also displayed in a similar manner. Separate plots are included for nighttime, twilight, and daytime measurements. Spring equinox results are presented that cover parts of the 23rd and 24th solar cycles from 2002 through 2010.

HEALTH, PHYSICAL EDUCATION, AND RECREATION

Accessibility of Outpatient Services as Viewed by Those Being Discharged from State Hospital South

Nicholas Baldwin and Jaren Blake

Osteopathic Medical Student III and Principal Investigator

HEALTH, PHYSICAL EDUCATION, AND RECREATION

Comparative Analysis of Prioritizing Professional Behavior between Undergraduate Students of Exercise Science and Future Employers

Bret H. Boyer, Paige Denney, M. Bohne, and Jason Slack

Utah Valley University

Students often put a lot of time and energy in professional behaviors that are different than what graduate programs and employers are seeking. The purpose of this study was to compare the perceived priority of professional behaviors between general Utah Valley University students, Exercise Science students, and employers in exercise science. A total of 572 students were surveyed to rank a randomized list of professional behaviors that were statistically compared with those of potential employers of this same student population. Analysis reveals that other than the #1 ranking, the classes differed from each other in ranking professional behavior and they differed from employers. Having identified these differences, we might consider designing academic programs to bring the students more in line with employers' priorities and doing it earlier in the students' academic careers.

HEALTH, PHYSICAL EDUCATION, AND RECREATION

Eat to Live, Live to Dance: Preventing Eating Disorders in Dancers

Kristy Jo Hunt

Dixie State College of Utah

Dancers often succumb to both internal and external pressures to maintain an ultra-thin appearance. Studies indicate this pressure results in 50% of dancers struggling with eating disorders such as anorexia and bulimia. Although the past few decades' research further identifies the problem, few preventative strategies have been implemented. Without proper education when challenged with inevitable negative body image, young dancers resort to their own means to change their bodies—unbalanced diets, self-starvation, and purging food right after eating—to attain their desired aesthetic appearance. I propose that a proper nutrition model be taught in all dance instructional facilities, both public and private, for dancers aged 9 to 18 years old.

HEALTH, PHYSICAL EDUCATION, AND RECREATION

Motivational Behaviors of Elite Senior Athletes

M. Vinson Miner and Bret H. Boyer

Utah Valley University

Qualitative research methods were utilized to analyze, compare, and develop insight into just how extrinsic motivation can affect intrinsic motivation. Extrinsic motivation moves from being more external in nature to a point where motivation is identified, integrated, or aligned with intrinsic motivation. An external factor could affect and help explain why elite senior athletes are motivated to engage in, participate, and compete in senior games and competitions. A ten-point survey/questionnaire was designed to illuminate the unique characteristics and commonalities associated with the participants. Qualitative components were observed, analyzed, and documented. (External rewards enhance motivation 77.5% responded in the affirmative). A variety of questioning techniques were employed including a Likert scale and fill-in-the-blank/open-ended questions. Participants' psychological responses, preferences, likes/dislikes, and attitude were tabulated and compared. In conclusion, the majority of respondents indicated that extrinsic rewards (on a scale of 1 to 10, average score was 7.58) were clearly important elements in their internalized or intrinsic motivation to engage, participate, and compete in Elite Senior World Games.

HEALTH, PHYSICAL EDUCATION, AND RECREATION

Patients' Opinions on Proper Reimbursement for Services Offered by their Primary Care Physicians

Travis Moulton

Osteopathic Medical Student III

HEALTH, PHYSICAL EDUCATION, AND RECREATION

Food Insecurities in Global Communities

Judith Pratt

Weber State University

HEALTH, PHYSICAL EDUCATION, AND RECREATION

Assessment of the Army's Method of Predicting Body Composition

Jason V. Slack and Jeffery Cowley

Utah Valley University

LETTERS—FOREIGN LANGUAGE, HUMANITIES, PHILOSOPHY

Phenomenological Intentionality of Pedro Salinas in his Poem “La memoria en las manos” from Largo lament

Andrew Bishop

Brigham Young University

Intentionality, in its various forms, connects the subject with objects as they appear within the subject's view of the world. Poets, like artists, create with their bodies and perceive the world with their senses and with their souls. Subjects allow objects to reveal themselves, to manifest themselves having identities according to the contexts in which they appear. This system is called intentionality—a phenomenological concept in which appearances have ontological meanings. This research focuses on philosophy as explained by Maurice Merleau-Ponty and sets up an interpretation of various poems by Pedro Salinas. “La memoria en las manos” exemplifies how the subject intends the stone and his hands while remembering an experience with the beloved. The poetic self in the poem probes the identities of objects in order to comprehend the essence of the beloved and of himself. The interpretation concludes the beloved is a conflation of Pedro Salinas's wife and lover, Margarita Bonmatí and Katherine Whitmore, respectively.

**LETTERS—FOREIGN LANGUAGE, HUMANITIES,
PHILOSOPHY**

**A Marriage Blessing: Masha, Pyotr, and a Tale of
Two Russias**

Michael Flynn

Dixie State College

**LETTERS—FOREIGN LANGUAGE, HUMANITIES,
PHILOSOPHY**

**What Holds the World Together at its Innermost?
The Correct Interpretation of Goethe's *Faust***

Eberhard Lehnhardt

Utah Valley University

Eleven days before Goethe died on March 11, 1832, Goethe said that *the greatness and ethical culture of Christianity as it glows and sparkles with splendor in the gospels* will never be surpassed by anything else. By employing comparison and contrast which Goethe used to create his works, one finds that *Faust* is Goethe's incommensurable call for a world steeped in absurdity and confusion to convert to the *greatest good*, the *New Covenant* of the gospels.

**LETTERS—FOREIGN LANGUAGE, HUMANITIES,
PHILOSOPHY**

**Oedipus's Family in José Luis Borau's *Furtivos*
and Icíar Bollain's *Flores de otro mundo***

Matías Martínez Abeijón

Southern Utah University

This article aims to compare the manner in which two films, *Furtivos* (1975) directed by José Luis Borau and *Flores de otro mundo* (1998) by Icíar Bollain, focus on the representation of a non-traditional family living in the Spanish countryside. Using as a departure point the metaphorical readings prevalent in the critical approach to *Furtivos*, I ex-

plore the meaning of the similarities between both plots in their representation of the family (i.e., the presence of a strong Oedipal tension that will come to be disturbed by the arrival of an outsider, the family as a microcosm for larger social and national issues) in the historical contexts in which these films were produced. I also analyze the historical significance of revisiting during the democracy the metaphors that had been previously used during General Francisco Franco's dictatorship. From a theoretical point of view, this discussion studies the presence of those clichés traditionally associated with film genre (i.e., melodrama's emphasis on the family table) while making reference to their relationship with the notions of gender and power.

LETTERS—FOREIGN LANGUAGE, HUMANITIES, PHILOSOPHY

Translation of Media as Education: Engaged Learning Project at Utah Valley University

Mariana Penin

Utah Valley University

LETTERS—FOREIGN LANGUAGE, HUMANITIES, PHILOSOPHY

Political Journalism and Pamphlets by Peruvian Women During the 1920s: The Role of Dora Mayer de Zulen

Iliana Portaro

University of California, Davis

In recent decades, feminist scholars in Latin American literature have challenged the canon by reevaluating authors whose work has been denied a place in literary history. This presentation examines the work of Dora Mayer (1868–1958), a Peruvian journalist and social activist, who, despite a limited access to education and professional training, managed to directly intervene in the traditionally masculine domain of politics and journalism. I analyze her coverage of the Cerro de Pasco Mining Company at the beginning of the 20th century, which depicted the deplorable working conditions of miners and the foreign exploita-

tion of resources. While her work and faith in social change through the press gained her immense recognition, the topics were seen as “inappropriate” for a female journalist and thus she was criticized and censored. Finally, I demonstrate Mayer’s strategies to challenge this notion that women were incapable of conveying the country’s political and economical reality.

LETTERS—FOREIGN LANGUAGE, HUMANITIES, PHILOSOPHY

Effects of the American Cold War Propaganda on Modern Perceptions of Russian Culture and Motives

Josh Uda

Utah Valley University

Americans are well aware that the Soviet Union strictly censored information and inundated its people with propaganda; however, most Americans are not aware that they were also heavily exposed to government propaganda during the Cold War. The results of this American propaganda include a persisting ignorance about the long and colorful history of Russia and abiding misperceptions about Russian culture and motives. This project compares predominant American misperceptions about Russian culture and history (as discovered through primary research conducted at Utah Valley University) with actual Russian culture and history (as portrayed by secondary research performed at Yale University). This project also examines examples of Soviet and American propaganda from the Cold War era and identifies causal relationships between propaganda-shaped misperceptions of Russia and discrepancies in modern media reports from American and Russian news outlets. Primary research for this project includes extensive interviews of U.S. citizens who have lived in Russia for a minimum of two years, interviews of U.S. citizens from diverse demographics who have never lived in Russia, a study of vintage, Cold War media, and a joint, Russian–American, headline-news translation project, funded by a CEL grant from the UVU Center for Engaged Learning. Of primary focus in this project is the Russo-Georgian War. Findings indicate that Russia’s actions in that conflict almost exactly mirrored U.S. actions in the first Iraq war; however, American news outlets initially portrayed the conflict as Russian aggression—an at-

tempted annexation of Georgia. Information about Georgian aggression and war crimes was suppressed in western outlets until The United Nations found Georgia at fault for the conflict.

LETTERS—FOREIGN LANGUAGE, HUMANITIES, PHILOSOPHY

The Project of Silencing—Why White Supremacist Capitalist Patriarchy Never Dies

Vanim Zetreus

Utah Valley University

Using Michel Foucault's analysis of "unreason" and "madness" throughout western history, it becomes possible to examine Hollywood images to show how uncomfortable presences of non-conformity are systematically silenced by being deemed insane. With that structure in mind, the psychoanalytic tradition of Sigmund Freud and Jacques Lacan provided insight to the implications of such a project of silencing. The picture painted is that of a large yet variously marginalized group being denied access to spiritual salvation, blocked from partaking of the discourse that hosts the "Good Old God," in Lacan's phrasing. Removal from the discourse denies access to God itself. The result underscores the hooksian concept of White Supremacist Capitalist Patriarchy as the position of privilege in society and exclusivity in terms of salvation.

LETTERS—LITERATURE

The Cold-Blooded Murder of the English Tongue: Addressing Grammar across the Curriculum

Julia Combs

Southern Utah University

Poor Professor Higgins. Instead of Eliza Doolittle, he should be taken out and "hanged" for the cold-blooded murder of the English tongue. Or should he? How important is it for college students to know the difference between *hanged* and *hung*? This presentation analyzes interviews with approximately 100 college professors across the curriculum

at Southern Utah University and Dixie State College. Many professors are collectively throwing their hands in the air (as well as pointing fingers toward English departments). Basic English skills should transfer across the disciplines. But critics, including those outside of higher education, have recently accused college students of being academically adrift when it comes to writing skills. In this presentation, I suggest that professors across the curriculum may find effective ideas to address basic grammar issues by communicating with each other, insisting that students write more, simplifying grammar “rules,” and trusting spoken language to help us. Peter Elbow’s “What Has Speech Got That Writing Needs” has good practical suggestions. Perhaps no one needs to be hanged. We all need to have common goals and communicate with each other, preferably in some form of acceptable English, about how we can meet those goals.

LETTERS—LITERATURE

Enamored of an Ass: A Shakespearean Exploration of Falling for the Wrong Guy

Jennifer Gibb

Northern Arizona University

Contemporary society is full of women who are willing to love, marry, and sacrifice for men who wreak havoc on their lives. This theme however is not a new one. Women throughout history have been willing to love men who are not good for them, and this is nowhere more evident than in the works of Shakespeare. Female characters in several of Shakespeare’s plays fall in love with men that do them harm physically, psychologically, emotionally, or all three. In *A Midsummer Night’s Dream*, Helena pines after Demetrius who verbally attacks and threatens her. Her devotion is evidence of the ideas of the culture and time period in which she learned to love. Titania’s love of both Oberon and Bottom reinforces idyllic view of love during the 16th century and demonstrates the patriarchal social structure. In *Othello*, Desdemona’s relationships with Othello and the other men in her life demonstrate the female role in society as well as social fear of female adultery and miscegenation. Ophelia in *Hamlet* is a stark example of the dire consequences that can befall wrong choice in love. Like Ophelia, Gertrude makes faulty amorous choices. Ophelia and Gertrude both emphasize the negative opinion of female sexuality in Shakespearean society, an idea that is still relevant today. An examination of these female charac-

ters' motives, the course of their relationships, and the consequences of their love in these three plays reveals to a great extent the nature of the female role in Shakespearean society as well as creates themes and situations that are still socially relevant to contemporary explorations of love.

LETTERS—LITERATURE

The Reign of Ants in Magical Realist Novels of Grass, Rushdie, and Marquez

Maria Inglefield

Weber State University

Magical Realism is a world literature, and its authors nod to each other in the building of the genre. One creature recurring in novels by Gunter Grass, Salman Rushdie, and Gabriel Garcia Marquez is the black ant. As ants make their way through the texts, they are used differently and for profound effect. Ants eat, ferry, swarm, march, and invade with the usual efficiency and stamina, and yet each author writes the ants as representative of something much darker than the reader expects, giving the insects fresh significance. The reader who at once remembers the ant as an inconsequential interloper pilfering from picnic baskets will be challenged with the portrayals and metaphors found in the novels *The Tin Drum*, *Midnight's Children*, and *One-Hundred Years of Solitude*.

LETTERS—LITERATURE

Destructive Modern 'Game of Chess' in T. S. Eliot's *The Waste Land*

Dustin Jackson

Dixie State College

This essay argues that "A Game of Chess" in *The Waste Land* carries the metaphor of rigid society through chess and the destruction of women in this game. The first form of criticism of T.S. Eliot's *The Waste Land* to comment on the themes of gender and sexuality was with James E. Miller, Jr., in *T.S. Eliot's Personal Waste Land* where he

claims the poem is an elegy for his same-sex lover Jean Verdenal. Miller even goes far enough to claim that events in the poem like meeting the woman in Germany in part one were activities Eliot and Verdenal engaged in together. This essay carries a certain standard; Eliot's events in his life are removed from the analysis. Women did not have the ability to control the game of chess during modernity. Any social norm, constructed around an "Absolute," is what Eliot argues against.

LETTERS—LITERATURE

The Museum, the Academy, and the Politics of Display: Fetishization and Aestheticization of American Indian Texts

Kimberli Lawson

Utah Valley University

With regard to the presentation of American Indian literature in the academy, there seem to be two common forms of display, which I will critique in order to tease out their shortcomings and offer alternative strategies for the presentation of American Indian literature. These two different but related forms of display are fetishization and aestheticization. After examining how these two forms of display function, I offer alternative pedagogical methods and philosophical lenses for the teaching of American Indian texts within the college classroom.

LETTERS—LITERATURE

Patriarchal Discourse in Charlotte Perkins Gilman's *The Yellow Wallpaper* and *If I Were a Man*

Kami McArthur

Dixie State College

Most people agree that for centuries women have been defined by their relationship to men, but few realize that women have in fact been defined *by* men. Through the use of language, women were persuaded to conform to domestic roles. In Charlotte Perkins Gilman's "The Yellow Wallpaper," the narrator is manipulated by her husband, John, to as-

sume a subjective, even child-like role. John's discourse with the narrator causes her to doubt herself and leads her to conform to John's perspective of who she is. Perkins again illustrates how a masculine-dominated language shapes female roles in "If I Were a Man." Here she provides a male-constructed definition of a "true woman." Even women who fit this definition are viewed as lesser than men. Although these stories were written over a century ago, the definition of a woman and her role is still heavily associated with a male-driven discourse today.

LETTERS—LITERATURE

The Function and Fetishism of Victorian Clothing in the Neo-Victorian Novel

Amanda Scott

Dixie State College

The Neo-Victorian novel, as it emerged in the latter half of the 20th century, is a dynamic genre as it not only redefines and rewrites the past, but also seeks to compare the modern world to the Victorian. One way the Neo-Victorian novel defines the Victorian and distinguishes it from modernity is through clothing. Victorian clothing is iconic for the period and represents rigidity and restriction, as well as the period's clearly defined gender roles. In A.S. Byatt's *Possession*, John Fowles's *The French Lieutenant's Woman*, and Neal Stephenson's *The Diamond Age*, clothing conveys a desire or nostalgia for Victorian culture and mores. By applying Valerie Steele's research on Victorian fashion and responding to other pertinent critical conversations, this essay argues that Neo-Victorian fiction fetishizes Victorian clothing to signify a longing for quaint and titillating sexual restriction and as metonymic of what we retrospectively project as the moral conservatism of the Victorian age.

LETTERS—LITERATURE

Robert Browning on the Unfaithful Woman

Tonia Wilson

Weber State University

Robert Browning's poetry evidences a decidedly individualistic approach to understanding women who commit adultery. Careful readings of *Porphyria's Lover*, *Andrea Del Sarto*, *My Last Duchess*, *A Forgiveness*, *The Worst of It*, and *The Statue and the Bust* reveal Browning's desire to portray unfaithful women as unique individuals in complex circumstances. This approach dissuades judgment by encouraging the reader's thoughtful evaluation before making assumptions. His sympathy toward the conventional, fallen woman is in stark contrast to the high-minded, social norms of the Victorian period. Also noteworthy is Browning's astonishingly accurate portrayal of such women; his insight into their mental workings is uncanny. By denying the Victorian urge to pass judgment on them, Browning's poems offer an unbiased, humanistic understanding of women who are unfaithful. This approach makes his work relevant well beyond the confines of the Victorian age.

PHYSICAL SCIENCES

Lorentzian and Gamma Distribution Representations of EFG Components in Inhomogeneous Broadening in PAC Spectroscopy

Mike Adams, P. Matheson, and W. E. Evenson

Utah Valley University

Perturbed angular correlation (PAC) spectroscopy is used to study the distribution and mobility of defects within crystals. The angular correlation of multiple gamma rays emitted from probe nuclei, affected by the net electric field gradient (EFG) in a probe's vicinity, are used to produce the PAC spectrum, $G_2(t)$. The distribution of EFGs from many random defects in a crystal results in inhomogeneous broadening (IHB) of $G_2(t)$. In cubic structures, distributions of two independent EFG components, which we identify as W_1 and W_2 , can be used to model IHB. The distributions W_1 and W_2 are functions of defect density. The distribution function for W_2 is well modeled by a single Lorentzian distribution, but the distribution for W_1 can only be characterized by a

sum of distribution functions, both of which are best fit by gamma distributions. This appears to be caused by the disproportionate contribution to W1 by defects in near and next-nearest neighbor sites. We report our progress in modeling W1 and W2 in cubic structures.

PHYSICAL SCIENCES

Who Can Perform an Observation?

Merrill Asp and Jean-Francois Van Huele

Brigham Young University

Because of the critical role of observation in quantum mechanics, we ask the question, "Who can perform a quantum observation?" We consider the possibilities of quantum systems observing each other by analyzing the tensor product of the two spaces of interest, i.e., spin space and momentum space, for the two systems. Defining observation as the correlation between characteristics in the individual spaces, we obtain a necessary condition for a given process (such as the measurement of spin in a Stern-Gerlach apparatus) to be accomplishable by the action of a quantum observable, namely the change of state described by the measurement postulate.

PHYSICAL SCIENCES

A Quantitative Analysis of the Water Quality of Coal Creek in Cedar City, Utah

Shannon Force, Samantha Huntsman Darren Dillard, Nicole Ramos, Kasidy Miller, Kim Weaver Ty Redd

Southern Utah University

Coal Creek, a small stream adjacent to Cedar City, Utah, has long suffered from a lack of plant, invertebrate, and vertebrate diversity. As the lack of invertebrate diversity is an indicator of poor water quality, our goal is to identify the factors contributing to this condition and their effect on the biological diversity in the creek. Heavy metals, runoff pollution, high sediment loading, and essential nutrient deficiencies are factors being investigated as any one of these alone can severely impair surface water quality. A variety of methods are in use, including chemical measurement techniques and biological assays. While this study

has yet to identify a single factor responsible for the diminished diversity in Coal Creek, preliminary evidence suggests that a combination of factors may be involved. Further analysis of annual floodwaters is anticipated to offer greater insight and provide more concrete conclusions.

PHYSICAL SCIENCES

Validation of Mineral Chelates by Scanning Electron Microscopy

Robert Jensen, Aldolph Yonkee, and Edward B Walker

Weber State University

Minerals often occur in nature as metal oxides or carbonates, which exhibit only limited solubility and are poorly absorbed from the digestive tract. To add value to minerals that are sold as dietary supplements, some manufacturers convert minerals from oxides into chelates. Such mineral chelates are composed of the metal ion and organic molecules that surround and bind to the mineral. This conversion process involves dissolution of the reactants in water, reaction, drying, and subsequent processing, all of which add cost to the final product. Some manufacturers skip this important step and simply mix dry chemicals in the supplement; however, dry-mixtures are not true chelates and should not be labeled as such. We have utilized scanning electron microscopy (SEM) in conjunction with elemental X-ray emissions to map images of simple blends and mineral chelates. Our resulting data and images clearly show the difference between “dry mixes” and true chelates. Such SEM-XRF image mapping provides a powerful new tool in the validation of mineral chelates.

PHYSICAL SCIENCES

Direct Determination of Iron in Whole Blood by X-ray Fluorescence

Shenelle Kleyn, Shalease Adams, R. Devin Roane, Firas Harb, Yasmen Simonian, and Edward B Walker

Weber State University

Iron content in blood is an important diagnostic tool and is among the most commonly performed tests in modern medicine. Currently, the majority of iron in blood is not measured directly, but is determined indirectly, either by counting red blood cells or by colorimetric determination of total hemoglobin. This study directly determines the actual concentration of iron in the blood utilizing X-ray fluorescence (XRF), regardless of its oxidation state or location. Samples of blood ranging across clinically relevant concentrations of iron will be measured in the ranges of 40–180 ppm, which corresponds to 5–20g/dL of hemoglobin. Removal of red blood cells by centrifugation allows similar XRF analysis of only serum iron. The values from XRF are compared with the more traditional clinical testing methods as well as inductively coupled plasma optical emission methods. We anticipate that this rapid analytical method for testing total iron in whole blood will add a new diagnostic tool to assist in clinical diagnoses.

PHYSICAL SCIENCES

Measuring the Change in EUV Reflectance Caused by Amorphous Silicon Oxidation

Zephne Larsen, David D. Allred, and R. Steven Turley

Brigham Young University

BYU Thin Films is trying to determine optical constants, specifically the index of refraction, of optics that reflect in the extreme ultra violet (EUV) light range (1–100 nm). These EUV mirrors are usually metal sputtered on a silicon wafer. Preliminary reports show amorphous silicon from the wafer oxidizes and affects the reflectance of the mirror. This talk explains how the continual oxidation affects the optical properties of EUV mirrors.

PHYSICAL SCIENCES**Characterization and Separation Applications of a Resorcinarene-based Cavitand****Na Li, Roger G. Harrison, and John D. Lamb.***Brigham Young University*

Resorcinarenes are cyclic tetramers that are synthesized from the condensation of resorcinol and various aldehydes. The phenolic hydroxyl groups on the upper rim of resorcinarenes can be modified easily to provide a preorganized host cavity. In our work, the upper rim of the resorcinarene is elongated by 4 alanine substituent groups. This deepened cavitand has several advantages. First, the deepened cavitand makes the capture of large molecules possible; second, the chiral alanine moieties on the upper rim provide stereoselectivity; third, 4 carboxylic acid groups are concentrated on the upper rim, which can enhance interaction with guests by non-covalent forces. Furthermore, these carboxylic groups can act as cation-exchangers for cation separations in ion chromatography. Here, we report the synthesis, characterization, and preliminary separation applications of this deepened cavitand.

PHYSICAL SCIENCES**Analysis of Dimethicone in Skin Protection Products by FTIR****Chase S. Naisbitt, Stuart A. Zuniga, and Edward B. Walker***Weber State University*

Dimethicone is a mixture of fully methylated linear siloxane polymers of various molecular weights. This water-insoluble material provides unique skin protection and lubrication when incorporated into skin care products. In fact, at concentrations of 1–30%, the U.S. Food and Drug Administration allows certain skin protection drug claims as per 21CFR347. Hence, it is important to have dependable methods of analysis for dimethicone in a variety of matrices. Current analytical methods required extensive sample preparation such as ashing and dissolution in acid prior to titration and are prone to experimental errors. We will present a new, rapid analytical method based upon Fourier Transform Infrared Spectrometry, utilizing an attenuated total

reflectance cell to determine concentrations of dimethicone directly in skin lotions and creams, without any sample preparation. This simplified method minimizes sample preparation errors and offers accurate results across a wide range of concentrations.

PHYSICAL SCIENCES

Modeling a History of Radiative Cooling of Various Size Planets and Planetesimals Using Maple

Alexander M. Panin

Utah Valley University

Maple is an advanced software package that can solve multidimensional partial differential equations (PDEs), yet with versatile graphic interface to illustrate the solution. However, an attempt to use Maple to solve a problem of radiative heating and cooling of bodies of finite thermoconductivity fails, due to requirement to solve linear heat diffusion PDE with the nonlinear boundary condition (BC). Indeed, a black-body radiation with intensity being proportional to the 4th power of surface temperature is a nonlinear BC. A canonic algorithm for solving the heat diffusion equation is based on the separation of spatial and temporal variables, which precludes the use of nonlinear boundary conditions. However, by modifying heat diffusion equation and intentionally making it nonlinear, it is possible to force Maple to use a different algorithm, which accepts some nonlinear BCs, including the black-body radiation condition. Reducing the nonlinear term to insignificantly small value for the problem at hand, allows accurate modeling of a wide variety of important radiative cooling/heating problems. We call this fruitful approach “delinearization” of linear PDEs. As an example, we model the dynamics of radiative cooling of some rocky planets and satellites of our Solar System after their formation. We also model the daily and yearly dynamics of the temperature distribution near their surface/subsurface. The details of our models and the results of our computation are presented.

PHYSICAL SCIENCES

Use of X-ray Fluorescence for Quality Screening of Minerals Used in Dietary Supplements

Brandon Price, Firas Harb, and Edward B Walker

Weber State University

Manufacturers of dietary supplements are required by the FDA to verify that product ingredients match the manufacturers' claims. One commonly used instrument for verifying mineral content in dietary supplements is the Inductively Coupled Plasma Optical Emissions Spectrometer (ICP-OES). One tool that might be used to potentially increase efficiency and costs of testing samples is a portable X-ray fluorescence (XRF) instrument. A Thermo Niton Xlt XRF was used to compare data with that of an Optima 2100 DV Perkin Elmer ICP-OES. Samples tested by the ICP-OES were also scanned with the portable XRF. The XRF used can effectively detect 10 of the elements most commonly used in dietary supplements. In addition to identifying elements the portable XRF can detect, a quantitative method for calcium was explored. The produced method showed results statistically comparable to the ICP-OES with a 95% confidence interval for concentrations of about 4–19%. Concentrations above this range no longer fit the generated calibration curve. It should also be noted that this method only applies to raw materials. Further restrictions include potential interelemental interferences from elements such as potassium and magnesium. It is recommended specific methods be explored for further elements.

PHYSICAL SCIENCES

Calculation of gravitational fields of disk-like and ring-like structures

Jeremy Redd and Alexander M. Panin

Utah Valley University

Calculation of gravitation field of disk-like and ring-like structures is of interest for astronomy (where some celestial bodies move in ring-like and disk-like structures gravitationally interacting with them), as well as for experiments probing gravitation on a small scale. There are some theories of gravitation which predict a deviation from inverse square law at a small (sub-mm) scale—and experiments at such small scale

have to focus on interaction of closely spaced discs (or even disk-like holes cut in solid discs) instead of classic Cavendish spheres (interaction of which at such distances is too weak to reliably measure). We model gravitation of disk-like and ring-like shapes by a classic straightforward procedure of slicing such shapes into finite elements and replacing those elements by point-like masses. However, calculation of gravitational field with high accuracy (especially in close vicinity or inside of such shapes) and with limited computing power requires serious deviation from such classic algorithms. In this presentation we discuss a few methods we used, their accuracy, advantages and disadvantages, as well as the results of computation of gravitational field of disc-like and ring-like structures of various sizes and aspect ratios both inside and outside of such structures.

PHYSICAL SCIENCES

Instantaneous Frequency Tracking Of Partial of a String Vibrating at Large Amplitudes Using a High-Speed Camera

David C. Ripplinger, Brian E. Anderson, and Timothy W. Leishman

Brigham Young University

Large amplitudes in a freely vibrating string cause an audible shift in normal frequencies (partials) of a piano tone caused by dynamic tension changes in the string. This phenomenon is often referred to as pitch glide. This paper presents experiments that allow extraction of pitch glide by measuring both instantaneous frequencies and total energy through the use of high-speed video camera recording. Experiments were performed using a repeatable plucking mechanism on a monochord string apparatus. The instantaneous total energy density at a point on the string is calculated as the string's motion decays to predict the pitch glide. The prediction is verified by the instantaneous frequency measurements.

PHYSICAL SCIENCES

Rapid Analysis of Zinc Oxide in Drug Products by X-ray Fluorescence

**Jazzi Sharifan, Trandon Bender, Kevin Larkin, Brian Penman,
and Edward B. Walker**

Weber State University

Zinc oxide is an active ingredient in a number of skin care products. Many of these products are registered with the FDA as over-the-counter drug products, thus requiring extensive routine quantitative testing of ZnO to insure quality and consistency. Currently, the most widely utilized test method for ZnO is based upon an EDTA titration. To avoid interferences from the product matrix, samples are currently ignited and then dissolved in acid prior to titration. This cumbersome, time-intensive process is subject to errors introduced during sample preparation. Advances in X-ray fluorescence technology have allowed this powerful analytical technique to be utilized in routine testing of creams and lotions with prescribed levels of zinc oxide. We have developed a simple, rapid analytical method that greatly simplifies ZnO analysis in such products, yielding results in less than one minute without the need for any sample preparation. We will report our procedure and present data demonstrating the advantages of this new method.

PHYSICAL SCIENCES

Accretion Disk Structure and the Fe K α Spectral Line Profiles in Active Galactic Nuclei

Brandon K. Wiggins and Brent A. Sorensen

Southern Utah University

Despite the attention that active galactic nuclei (AGN) have received in astrophysics and observational astronomy, the structure of their broad line regions continues to be an area of active research. An accretion disk in these central, strongly emitting regions has been inferred in some AGN on account of their doubled-peaked Fe K α spectral line. In this preliminary study, we utilize simple computer models of accretion disks to investigate approximate Fe K α line profiles resulting from various disk configurations and structures. Pertinent physics regarding AGN is briefly discussed. We compare our results with observation and

the literature and provide discussion on the possibility of various disk configurations and phenomena in the context of current AGN theory.

PHYSICAL SCIENCES

Romancing Mathematics With Chemistry—How Mathematical Trees Can Be Used to Synthesize Molecular Structures

Chin-yah Yeh

Utah Valley University

Structures of chemical compounds can be synthesized and categorized through mathematical means. Organic compounds are suitable for this technique because of their simple valences. Mathematical variables serve as chemical symbols and mathematical equations are chemical structure generators. Five classes of acyclic organic compounds are generated: I. Compounds made of hydrogen and second row elements, C, N, O, and F; II. Category I compounds with unstable peroxides and hydrazo compounds excluded; III. Category II compounds with relatively unstable geminal polyols and polyamines excluded (this category is the norm of organic compounds); IV. Unsaturated compounds; and V. Aldehydes and ketones.

POSTER SESSION

The Effect of a Regional Exercise Science Professional Conference on Student Perception of Professional Behavior

Paige Denney and Bret Boyer

Utah Valley University

POSTER SESSION

Analysis of Barn Owl Pellets (*Tyto alba*) Collected from Davis County, Utah

Jaime S. Heiner and Stephen P. Niedzwiecki

Oquirrh Mountain Charter School.

Thirty pellets from common barn owls (*Tyto alba*) were collected on November 15, 2010, from a known roosting site in Kaysville, Utah. This site provided an interesting combination of habitats, in that the core of the barn owl's range is largely agricultural while the periphery consists of rapidly developing suburbs. Faunal remains represented within the pellets were identified at least to genus level using dichotomous keys and the osteological reference collection at the Utah Museum of Natural History. Preliminary results indicate that *Microtus* sp. represent the majority of the assemblage; however, bats and birds are also represented. Although the barn owl bias for *Microtus* sp. is well-documented, we feel this study provides significant baseline data for examining the loss of biodiversity that is associated with the rapid urbanization prevalent in northern Utah.

POSTER SESSION

Defense Chemistry in *Inga umbellifera*

Susanna Khachatryan, Thomas A. Kursar, Phyllis D. Coley, and John Lokvam

University of Utah

Over half of the macro-organisms in tropical forests are plants and their herbivores. Because of their tight co-evolutionary history with herbivores, tropical plants produce chemical defenses to protect vulnerable tissues such as their leaves. In tropical forests, over 70% of a leaf's lifetime damage occurs in the first few weeks of expansion; thus, young leaves are under a great deal of selective pressure to defend themselves. Utilizing qualitative and quantitative methods we analyzed defense chemicals in young leaves from the widespread species *Inga umbellifera*, collected from 5 sites across South and Central America. By extracting and separating soluble leaf components, further separating

them using high-performance liquid chromatography and determining molecular weights by mass spectrometry, compound identification was achieved using a defense chemical database constructed in our lab. These analyses revealed that a variety of defense chemotypes could be observed over the distribution of this species. This suggests that within a species, differential selective pressures can result in altered gene expressions. Intraspecific analyses such as these can provide insight into how a plant's defense chemistry varies across its geographic range. Identifying the amount of chemical defense variation among individuals within a species helps in identifying the degree of plasticity in defense chemical expression, given similar genetic makeup. Since plants and their herbivores make up most of the macro-organisms in tropical forests, understanding the dynamics of such interactions is important. Unraveling the intraspecific plasticity in expression of chemicals that plants produce is part of an effort to better understand the widespread nature of plant-herbivore interactions.

POSTER SESSION

A Comparative Analysis of Increased Childhood Obesity and the Reduction of Physical Education Programs in Public Education in Utah Public Schools

Kemp Rosqvist and Bret Boyer

Utah Valley University

POSTER SESSION

Digital Humanities: What's All the Buzz About?

Burke Sorenson

Utah Valley University

Across the country, digital humanities, sometimes also known as humanities computing, has taken higher education by storm. But few outside of the Humanities discipline and many within it do not understand its methodological nature and interdisciplinary scope. While academic departments of the digital humanities typically include technical practitioners as well as traditionally trained scholars with experience or ex-

pertise in digital media, finding qualified digital humanities professors remains a challenge. The goal of many researchers in the digital humanities is to begin to integrate technology into their scholarly activities, such as the use of text-analytic techniques; GIS; commons-based peer collaboration; interactive games and multimedia in the research and teaching of history, philosophy, literature, religious studies, or sociology. This presentation will look at digital humanities including electronic publication and computational analysis in project design and visualization, as well as in data archiving and retrieval.

POSTER SESSION

How Ecological Science Is Portrayed in Mass Media

Leilani F. Williams, Jerald B. Johnson,
Brigham Young University

Matthew J. Baker
Northern Kentucky University

Although scientists publish thousands of peer-reviewed journal articles each year, only a small percentage of the findings reported in these articles find an audience outside of academia. Studies that do find a broader audience rely primarily on mass media to convey research results to the broader public. However, there is a perception among scientists that research findings presented in mass media selectively represent the results and often fail to convey essential elements of the work. We examined the validity of this perception by examining how ecological science has been portrayed in the mass media. We surveyed articles published in the journal *Ecology* over the past decade to answer four basic questions: (1) what proportion of ecological research is picked up by mainstream media; (2) which type of media is most likely to cover primary ecological work; (3) which components of published work are most commonly highlighted in the media; and (4) is media coverage of ecological research increasing, decreasing, or static? Our results show that limited article coverage and selective inclusion of content are common by mainstream media and could affect the way that the general public perceives scientific research.

SOCIAL SCIENCE

Heathens, Neo-Pagans, and Non-Believers: A Look at the Attitudes of the Non-Affiliated at Salt Lake Community College

Karine Agajanian
University of Utah

Jen Chapman and Spencer H. Blake
Salt Lake Community College

In the past year, two major studies on youth and religious affiliation have generated interest not only within the academic world but also in the public sphere. Robert D. Putnam and David E. Campbell's work "American Grace: How Religion Divides and Unites Us" and Drew Dyck's "Generation EX-Christian: Why Young Adults are Leaving the Faith...And How to Bring Them Back" indicate a trend: The youth are leaving the churches of their heritage at a higher rate than previous generations. This study is a small sample that looks at the religious attitudes of the non-affiliated students at Salt Lake Community College and compares these attitudes to Dyck's categories for the non-affiliated.

SOCIAL SCIENCE

The Impact and Interdiscursive Mechanics of Globalization on a Cambodian-American Refugee

Brandon J. Bales
Brigham Young University

Cambodian refugees in the processes of resettlement or emplacement in the U.S. occupy a dual position of patient and agent across several scales of globalization phenomena. Analyzing narratives told by Cambodians in Utah about their refugee experience reveals how the process of resettlement operates through narrative construction and performance. Refugee narratives draw from and coordinate a diverse amount of cultural and global information. In review of globalization and refugee resettlement literature, this paper first seeks to establish clear definitions for and interrelationships between macro and micro scales of globalization and the conscious experience of these scales at the individual level. This study examines personal narratives as metapragmatic

constructs and commentary about how global processes shape the personal identity and social integration of Cambodian refugees in the U.S. Through metapragmatics, refugees create narratives as classic Bakhtinian voicing structures that organize diverse and, sometimes, conflicting elements of cultural meaning into a unified and coherent text. With the purpose of resolving the upheaval from genocide and migration to foreign circumstances, the telling of these coherent texts is a ritual performance that establishes a coherent reality. This paper introduces the concept of dicentization as a mechanism for constructing cause and effect relationships. It also discusses the specific semiotic mechanisms that combine diverse information into an interdiscursive text.

SOCIAL SCIENCE

The Emergence of Leadership in Online Gaming

Chris Becker

Weber State University

SOCIAL SCIENCE

The Temple Recommend: A Solution to the Free-Rider Problem

Austin Bowles and Tyler Bowles

Utah State University

Temple worship and geographical boundaries are unique features of the Church of Jesus Christ of Latter-day Saints (LDS Church). Our research uses economic theory to understand how the organizational benefits of the temple recommend system and geographical ward boundaries may solve the market failures associated with religious organizations. These features of the LDS Church will be used to create a framework for understanding how the church limits free-riding and sub-optimal participation, which would result in increased efficiency and production of large religious benefits. Because the market failures faced by the church are the same as those faced by other religious groups and similar to those faced by other club-like organizations, this research provides insights into possible solutions for groups besides the LDS Church.

SOCIAL SCIENCE

A Second Soul: Exploring Contexts of Language Acquisition and Personality

Aaron K. Combs

Dixie State College

Charlemagne said, “To have another language is to possess a second soul.” Language plays an important role in personality development. It may also aid in the development of a second personality, or “a second soul.” Through classical conditioning processes, language can prime relatively stable patterns of behaviors and feelings. Dramatic changes in an individual’s language, such as learning a new language, can set up new stable patterns of communication and personality. In this paper, I present research on classical conditioning principles relevant to language and personality. I also present primary research, including survey and case studies demonstrating the effects of language on personality. Individuals who learn and speak a new language while performing social activities often experience a personality change. For example, as LDS missionaries proselyte in a new language, they become classically conditioned to feeling extroverted in social exchanges in that language. This paper indicates that this classical conditioning does not generalize to their native language.

SOCIAL SCIENCE

Learning About Peace and Justice Work

Josh Gold

Salt Lake Community College

Extensive interviews were conducted during 2008–2009 for the purpose of designing and planning a Peace and Justice Studies (PJS) Curriculum for the School of Humanities and Social Sciences at Salt Lake Community College. Here, I present some of the insights gleaned from professionals whose work in public and civil society organizations in the Salt Lake Valley deliberately promotes peace and justice locally, nationally, and globally. As a result of learning about peace and justice work in the community from them, I propose that such civil society links should help people across Utah address the gender, racial, ethnic, class, and ideological divides that cause conflict and costly divisions in

our community. The need for a PJS program to work with students in the development of a cost accounting—of waging peace vs. war in the 21st century—became apparent and, as a result, 2 service-learning Peace and Conflict courses (1 domestic and 1 international) were developed as the anchor of the curriculum. In addition, the sabbatical time and the 77 subjects interviewed helped me to construct a model of how violence operates, which I call “A Political Economy of Direct Violence.” It became another way to teach for peace and justice.

SOCIAL SCIENCE

Returning to School: Women at Salt Lake Community College

Anne Graham

Salt Lake Community College

SOCIAL SCIENCE

Six Percent Policing: Focusing on Chronic Offenders

John Hill

Salt Lake Community College

Historically, law enforcement practitioners would lament that a small percentage (“about 5%” as typically told) of the population created the vast majority of crime and disorder diminishing the quality of life for the remaining 95% of generally law-abiding citizens. More scientifically, it was found to be precisely 6% when Dr. Marvin Wolfgang published the results of his landmark study, “Delinquency in a Birth Cohort” (1972), in which he used official police records to follow a cohort of males born in Philadelphia in 1945. The study came to identify what were termed *chronic offenders*, who comprised only 6% of the total sample, but were responsible for over half of all offenses committed by the cohort. Subsequent and related/replicated research has also identified further evidence of chronic offenders, or “Six Percenters.” Here, the author recommends that agency resources be focused more precisely upon chronic offenders and for practitioners to embrace the new paradigm of *offender-focused* policing.

SOCIAL SCIENCE

The Regressives: Unraveling the Progressive Movement

John Howell

Southern Utah University

The Progressive Movement of the late 1800s and early 1900s was one of the most influential political movements in American history. Progressives sought to shift power away from political elites and into the hands of the citizenry. Progressives viewed more democratic government as both less corrupt and more responsive to the will and needs of the people. The early Progressives were very much in line with the Populist movement of the same time period. The modern political landscape has many who claim the "Progressive" label, yet have abandoned their belief in the tenets of the Progressive Movement. The goals of modern "Progressives" are more often pursued through the very elite powers that the original Progressives opposed, often in direct opposition to the stated wishes of "We the People." In effect, they have become that which their ideological ancestors despised.

SOCIAL SCIENCE

Changes in Racial Concentration Ratios in Mountain States Counties, 1990–2000

L. Dwight Israelsen

Utah State University

Studies of residential segregation have concentrated on racial/ethnic segregation in neighborhoods, standard metropolitan areas, counties, and states and have looked at socioeconomic status and other variables to explain differences in segregation measures across groups and over time. Commonly used segregation indices measure segregation of a particular race or ancestral group in an area by examining the distribution of that group across sub-areas. In a recent paper, the authors introduced a new measure of segregation, the racial concentration ratio (RCR). The RCR is a one-parameter measure of the racial (ethnic, ancestral) composition of a subdivision of an area relative to the racial composition of the entire area, and is similar to a Gini coefficient. In the previous study, the RCRs were based on the percentages of county

(state) population in six groups: Non-Hispanic Whites, Hispanics, Blacks, Asians (Asians/Pacific Islanders), American Indians (Indians and Alaskan Natives), and All Other, relative to the entire U.S. percentages. First, the percentage of each group in a county or state was scaled by the overall U.S. percentage of that group. Then, RCRs were calculated treating these cumulative scaled percentages as "cumulative income" is treated in the income Gini coefficient calculation (i.e., on the y axis of the Lorenz curve) and treating the fractions 1/6, 2/6, 3/6, 4/6, 5/6, and 6/6 as "cumulative population" is treated in the income Gini coefficient calculation (i.e., on the x axis of the Lorenz curve). Hence, if the racial composition of an area is exactly like that of the entire U.S., the RCR is zero, and as the racial composition increasingly diverges from that of the U.S., the RCR grows closer to one. In the current study, RCRs are calculated for all Mountain States counties for 1990 and for 2000 by scaling the racial/ethnic group percentage in the county by the overall Mountain States percentage, and RCRs were calculated accordingly. Hence, the county RCR shows how closely the racial/ethnic composition of the county mirrors that of the region. Changes from 1990 to 2000 in county RCRs are calculated, and an econometric model including economic, social, education, demographic, political, and location variables is developed and used to identify the determinants of changes in RCRs at the county level.

SOCIAL SCIENCE

Lessons Learned from C.R.I.M.E.S.

Ramona Linville Higley

Weber State University

During a two-year period from 2003 to 2005, students attending BYU-Idaho participated in an academic exchange of letters with several prisoners who were incarcerated in the maximum security section of the Utah State Prison system in Draper, Utah. This program was known as Correctional Inmates Educating Students (C.R.I.M.E.S.). The purpose of this experience was to give students intending careers in the criminal justice field the opportunity to interact with, ask questions of, and learn from inmates concerning crime and the criminal justice system. This paper summarizes some of the insights gained from these letters.

SOCIAL SCIENCE**Antigay Behaviors Among Young Adults:
Prevalence, Patterns, and Motivators in a
Noncriminal Population (Replicated)****Chad Morrow, CoCo James, and Spencer Blake***Salt Lake Community College*

This study is a replication of a study originally conducted in 1998 by Dr. Karen Franklin of Alliant International University. The abstract of Dr. Franklin's paper is quoted hereafter. "This is the first empirical research into prevalence rates of and motivations for antigay harassment and violence by noncriminal young adults. In an anonymous survey of 484 young adults, 1 in 10 admitted physical violence or threats against presumed homosexuals, and another 24% acknowledged name-calling. Factor analyses revealed four motivational themes: peer dynamics, antigay ideology, thrill-seeking, and perceived self-defense. Compared with nonassailants, assailants held more negative attitudes toward homosexuals and reported more negative social norms among their friends. Assailants also had higher levels of masculinity ideology and social drinking. The findings suggest that many young adults believe antigay harassment and violence is socially acceptable, particularly in response to inferred sexual innuendos or gender norms violations. Because antigay behaviors are culturally normative and usually go unreported, educational outreach to adolescents and preadolescents is likely to be a more effective prevention strategy than criminal prosecutions under special hate crimes laws." Both studies focused on rates of antigay attitudes and behaviors among community college students. A smaller percentage of respondents in Salt Lake City reported participation in incidents of antigay violence, but there were similar rates of name calling in both studies. Differences were discovered in demographic data (particularly rates of religious identity and racial demographics) and in the relationships between assailants and targeted homosexuals.

SOCIAL SCIENCE

The Influence of Presidential Subscription to Schools of Political Thought on the *Paradigm of Imperial Decline*

Brittney Parks

Southern Utah University

SOCIAL SCIENCE

Shared Promise: The Rhetoric of Presidents Bush and Obama

Luke Perry

Southern Utah University

The election of Barack Obama was viewed by many as a pivotal moment in American politics. Much was made of the contrast between President Obama and his predecessor, George W. Bush. Two years into the Obama administration, presidential scholars are beginning to re-think the perceived gap between these two presidents. Important differences undoubtedly remain, yet continuity has been a dominant theme in several domestic and international policy areas from the Bush tax cuts to the War on Terror. This paper will examine the presidential rhetoric employed by Presidents Bush and Obama during their respective campaigns and their first two years in office. In contrast to popular perceptions, both men employed similar rhetorical frameworks in seeking the presidency and seeking support for their desired policies.

SOCIAL SCIENCE

Demographic Transition and Religious Identity among Mormons in the Intermountain West, 1972–2010

Rick Phillips

University of North Florida

Sociologists have theorized that in regions where one religion dominates, members of the majority faith develop a rich religious subculture that enhances and sustains religious identity and church participation. Conversely, when churches must compete in a pluralistic setting, religious switching is common, and religion loses much of its public power. Within the U.S., no denomination has a larger a share of any state's population than does The Church of Jesus Christ of Latter-day Saints (the LDS or Mormon Church) in Utah. Scholars have argued that Mormonism fosters a distinct religious culture in Utah, and studies show that Mormons in Utah and surrounding environment have traditionally had higher rates of religious participation than their counterparts in other parts of the country. However, since 1989, the Mormon share of Utah's population has been steadily shrinking, and some counties that once had large Mormon majorities are now increasingly pluralistic. This paper uses data from the General Social Survey from 1972 to 2010 to show how changing demographics in Utah and the Intermountain West have impacted the religious culture and patterns of religious participation of Mormons in the region.

SOCIAL SCIENCE

A Test of Emerson's Network Exchange Theory

Emmalou Rees

Weber State University

SOCIAL SCIENCE

Influence of Mood Induction on Impulsive Decision Making

Katrina Rodzon, M.S. Berry, A. Odum, and K. Jordan

Utah State University

High levels of impulsivity are associated with several maladaptive behaviors including, but not limited to, drug abuse and addiction. Negative mood may affect the ability to maintain self-control, and evaluating practices that positively influence mood may help to identify techniques that increase self-control. This experiment aimed to investigate the role of mood induction—using a series of self-reflection statements (Velten statements)—on impulsivity as measured by a delay discounting measure. A between-subject design was employed in which participants saw (1) positive, (2) negative, or (3) neutral Velten statements and then completed a delay discounting procedure in which a choice was made between immediate and delayed hypothetical monetary outcomes. Specifically, participants were asked to make a decision between money now versus a larger amount of money after a delay (i.e., 1 week, 2 weeks, 1 month, 6 months, 1 year, 5 years, 25 years). Preliminary analyses demonstrate differences in level of impulsivity between the three mood induction groups with those in a negative mood displaying the most impulsive decision making (i.e., steeper discounting).

SOCIAL SCIENCE

Internal and External Threats Posed by Islamic State and Sub State Actors

Maxwell Sands

University of Utah

Pan-Islamism has been a much-sought-after goal for almost every generation since the death of Mohammed in 632 AD. Even though some of the greatest Islamic figures and Qur'anic scholars have attempted to create a unity within the Islamic community, nothing has been able to heal the wounds created nearly 1400 years ago. In Samuel Huntington's 1993 article, "The Clash of Civilizations," his fear-mongering rhetoric displays Islam as one homogenous unit. Even in towns in the U.S. with little religious diversity, it is possible to observe the difference in be-

liefs in Islam (e.g., Salt Lake City, Utah. has 3 mosques, each with its own benefits and ideological uniqueness). It is now necessary for a new generation of political scientists with an intricate understanding of Islam to emerge from the ashes of Huntington's archaic views and approach modern conflict and international relations through a prism of understanding and not fear. In doing so, political discourse and strategy can be pro-active and not just active as it has been in the U.S.

SOCIAL SCIENCE

The Significance of Physical Beauty: Acid Attacks and the Media's Message to Women

Shaylee Wheeler

Weber State University

SOCIAL SCIENCE

Urban Legends in Japan and the United States: A Cross-Cultural Analysis

Mai Yamamoto

Weber State University