CONTENTS

Section I: Arts and Humanities

4 ARCHITECTURAL THEORY: Jonathan Boelkins. The Vietman Memorial

10 MUSIC HISTORY: Erin Brooks. The Dies Irae and Totentanz

14 ENGLISH: Shauna Copeland. Double Victims

22 ART HISTORY: Laurel Eddleman. Gerard David’s Nativity Triptych

29 LANDSCAPE ARCHITECTUE: Juana Gregory. Van Buren Riverfront

36 ARCHITECTURE: Melissa Harlan. Dar Islam Mosque

NOTE: See the Inquiry website for the Brooks and Copeland complete articles.

Section II: Social Sciences and Business

44 ECONOMICS: Maria Cadario. Free Trade: A Case Study of Brazil

56 FINANCE: Stephanie Gosnell. Intrastate Variation in Higher Education Funding

63 PSYCHOLOGY: Eric Jackson. Organizing Madness: Nosology in Historical Perspective

69 HISTORY: Karilyn Moeller. Chinese Women Unbound

NOTE: See the Inquiry website for the Cadario, Gosnell, and Jackson complete articles.

Section III: Sciences and Engineering

80 BIOLOGY: Lindsey Davis. Transcriptural Control of the Opp Operon

88 PLANT PATHOLOGY: Sarah Doege. Natural Calcium Oxalate Crystals

95 MATHEMATICS: Andrew King. Mathematical Interpretation of Political Power

101 MECHANICAL ENGINEERING: Robin Prince. Nanoindentation and Tip Geometry

106 HORTICULTURE: Alisha Sanny. Blackberry Cultivars

NOTE: See the Inquiry website for abstracts of papers considered for publication but not received by the publication deadline and for the following papers recommended for publication by the Inquiry Publications Board Subcommittees, but which had to be excluded because of space and cost: FOOD SCIENCE. Mandy Cox. Wound Inducible Genes Encoding Enzymes; CROP, SOIL AND ENVIRONMENTAL SCIENCE. Bill Hendrix. Anti-Fungal Defense Responses in Tobacco; PHYSICS. Brian Sawyer. Multiphoton Interaction; COMPUTER SCIENCE/COMPUTER ENGINEERING. Thao Doan. Improvement of Latent Semantic Analysis Performance; CHEMISTRY. Justin Thompson. Liquid Water on the Martian Surface; and CROP, SOIL AND ENVIRONMENTAL SCIENCE. Rupali Ugrankar. DNA Sequence of Melanocortin 1-Receptor Gene in Coturnix Japonica. Website address:

http://advancement.uark.edu/pubs/inquiry
Foreword

This journal, the fourth in an annual series, is a project of the Teaching Academy of the University of Arkansas and is testimony to the Academy's belief that a function of good teaching is to encourage good research and creative thinking on the part of the students.

This issue of Inquiry records the individual research exploration of fourteen U of A student/faculty mentor pairs during the 2002/2003 academic year, plus one project, Laura Eddleman's, that was completed during the previous year but for which permissions to use her illustrations could not be secured by publication time for the previous year's edition. The projects included here are drawn from disciplines from five of the university's six undergraduate colleges and schools—the Dale Bumpers College of Agricultural, Food and Life Sciences, the School of Architecture, the J. William Fulbright College of Arts and Sciences, the Sam M. Walton College of Business Administration, and the College of Engineering—and are representative of the quality of research done by the honor students in the various disciplines represented on campus. The breadth of subject matter included here is testimony to the commitment made throughout the university to honors study and research at the undergraduate as well as the graduate level. These fifteen articles were chosen by Inquiry's publication board from abstracts received as a result of a call for papers. They vary in subject, in writing style, and in the manner in which they reference their research sources; but they are uniformly excellent in content. In each case, the paper published herein is a précis of the student's larger research product. Many more abstracts were received than could be accepted for publication. This year for the first time the journal is paired with a website, http://advancement.uark.edu/pubs/inquiry/, on which a number of abstracts not selected for publication but viewed as particularly interesting by the publications board subcommittees, as well as six completed articles recommended for publication but excluded because of space and cost limitations. We hope interested readers will peruse both the articles in hard copy in the journal and those published on the website as well.

The intent of the journal is to record the depth and breadth of the scholarly activities of the university's best undergraduate students. I believe that it does this.

Murray Smart, Jr., Editor, University professor of Architecture, Emeritus

Inquiry Publication Board, 2002-2003 Academic Year

Subcommittee I: Arts and Humanities. Mark Cory, Foreign Languages, chair.

Mark Boyer, Landscape Architecture; Ethel Goodstein, Architecture; Amy Herzberg, Drama; Suzanne McCray, Honors; Louise Montgomery, Journalism; Lyna Lee Montgomery, English; Ken Stout, Art; Janice Yoes, Music.


Phil Besonen, Curriculum and Instruction; Chuck Briton, Economics; Lynda Coon, History; Paul Cronan, Computer Information Systems and Quantitative Analysis; Sue Martin, Human and Environmental Sciences; John Norwood, Accounting; Mary Jo Schneider, Anthropology; David Schroeder, Psychology; John Todd, Management.

Subcommittee III: Natural Sciences and Engineering. John Hehr, Geosciences, Chair.

Neil Allison, Chemistry; John Clark, Horticulture; Allan Cochran, Mathematics; Ro DiBrezzo, Health Science, Kinesiology, Recreation, and Dance; Bob Elliott, Civil Engineering; Bill Harter, Physics; D. M. Ivey, Biology, Max Meisch, Entomology; Marianne Neighbors, Nursing.
SECTION I: ARTS AND HUMANITIES

ARCHITECTURAL THEORY, MUSIC, ENGLISH, ART HISTORY, LANDSCAPE ARCHITECTURE, AND ARCHITECTURAL DESIGN
THE VIETNAM MEMORIAL:
A POSTMODERN REFLECTION

By Jonathan Boelkins
Department of Architecture

Faculty Mentor: Assistant Professor Korydon H. Smith
Department of Architecture

Abstract:

Complex in its cultural significance and entanglements, the Vietnam War is an event that continues to reverberate with social dissonance. The Vietnam Memorial (Maya Lin) sustains multiple, often oppositional, debates surrounding the Vietnam War. The Vietnam Memorial: A Postmodern Reflection examines the significance of the memorial with regard to American cultural history. The physical experience of the memorial—the decisive yet subtle geometry, the polished black marble, the chronological listing of names, and the scarring of the ground plane—is described with particular focus on the bond between material and social “reflection.” This paper utilizes the Vietnam Memorial to discuss architecture’s ability to evoke, through sensuality, a state of meditation that allows an individual to contemplate his/her relationship to social history. The author states: “Ultimately, the Vietnam Memorial becomes a reflection; a reflection not only of the [literal] image of the visitor, but of thoughts, feelings, and beliefs; visitors are allowed to see themselves and see inside themselves.” Architecture’s role in allowing, rather than denying, multiplicity and conflicting social ideologies is discussed.

Despite modernism’s attempt to create a universal architecture and a singular order, subsequent critiques recognize the limitations of such a prescribed response. By pursuing a single ideology at the expense of all others, modernism created works that rejected multiplicity and openness. The question becomes whether or not architecture can be crafted in order to accept the imperfections and discordant expectations of its users, rather than impose a single unifying thread that forces one order throughout.

The Vietnam Memorial offers a potent illustration of the power of consciously crafted multiplicity by refusing to frame a single ideology or historical perspective. Consequently, an architectural solution developed where a sociological solution seemed impossible. Marco Frascari writes in support of architecture’s unique capacity for resolution by saying, “in the details are the possibilities of innovation and invention, and it is through these that architects can give harmony to the most uncommon and difficult or disorderly environment generated by a culture.” While the Vietnam Memorial does not craft details in the literal manner that Frascari describes, the spirit of spacemaking and the pursuit of a connection between user and place resonates with Frascari and becomes a powerful illustration of architectural resolution. As Juhani Pallasmaa recognizes,

“Architecture is the art of reconciliation between ourselves and the world, and this mediation takes place through the senses.” The monument uses the potential of architecture as mediator to distance itself from the conflict. By consciously avoiding traditional political or historical stances, the wall acts a conscientious objector, choosing instead to engage the visitor’s senses, encouraging reflection and interaction.

In contrast, the Marine Corps War Memorial epitomizes the traditional war memorial. The sculpture of soldiers hoisting a flag over hard-fought enemy territory makes no mention of loss, only of victory. Furthermore, the statues are idealized in form in their representation of the Pulitzer Prize-winning image taken by Joe Rosenthal. The soldiers hoisting the flag stand
considerably larger than life, and the memorial does not list the individual names of the soldiers who died in the battle for Iwo Jima, or any other battle.

Perception becomes the fundamental difference between the two memorials. Very few questioned the reason that the war in the Pacific was being fought, and even though the invasion of Japan never happened (which was a primary reason for taking Iwo Jima,) no one ever considered the loss of life on Iwo Jima as meaningless. Americans were galvanized by the image of the American flag flying over formerly enemy territory. Such an image was never produced in Vietnam. Instead we are left with the images of street fighting and napalm raids. Similar images were produced in World War Two, but were not released to the public until after the war, as the government clearly understood the effect such images had on national unity and morale. The Vietnam War was the first war broadcast daily into American life, becoming known as “the living room war.” The immediate reporting of graphic depictions of combat helped to create stark divisions in American society.

In the words of Jan Scruggs, “A Vietnam Veterans Memorial bore a special burden. It had to satisfy audiences with conflicting needs and expectations.” The “conflicting needs” ranged from proponents of the war to anti-war protesters, the families and friends of those killed, and the veterans themselves, to name but a few. A polarized society found an opportunity for reconciliation, but not necessarily healing, through the architectural genius of the Vietnam Memorial, which accepted the reactions of visitors, encouraging them only to consider the loss through each name.

However, openness at the level of interaction cannot be interpreted as ambiguity in the design phase. Engineering experiential openness often requires greater control than a traditional closed design. While the memorial allows a variety of responses, its multiplicity is independent from indeterminacy. The wall observes a strict geometry, so that while the cultural, historical and political landscapes are recessed into that geometry, the surface of the wall is free to be manipulated in phenomenological ways.

In Umberto Eco’s essay, “The Poetics of the Open Work,” composer Henri Pousseur describes his composition Scambi by saying, “Scambi is not so much a musical composition as a field of possibilities, an explicit invitation to exercise choice. It is made up of sixteen sections. Each of these can be linked to any two others, without weakening the logical continuity of the musical process.” The critical link between Scambi and the Vietnam Memorial is Pousseur’s descriptive phrase, “an explicit invitation to exercise choice.” Pousseur thus makes the link between a deterministic approach of the composer and the less deterministic process of the musician. Understandably, the musician is more analogous to the builder than to the user, but the important issue here is recognizing the clear and deterministic intent of the creator. Maya Lin states, “I never expected it to be passionless. The piece was built as a very psychological memorial. It’s not meant to be cheerful or happy, but to bring out in people the realization of loss and a cathartic healing process.”

The phenomenological and haptic properties of the Vietnam Memorial that make it open were by no means left to chance. According to Pallasmaa, “an architectural work is great precisely because of the oppositional and contradictory intentions and allusions it succeeds in fusing together.” In what seems at first glance to be a simple gesture of remembrance are layers of meaning and tectonic expression coalescing to form a moving tribute. Robert Campbell writes, “The wall is a huge book, open at a place where it both begins and ends, and its text, its long march of names has made it, you realize, a memorial to individual human beings rather than to any larger but vague concept of country or sacrifice or victory or heroism.” By simply presenting the names of those killed rather than the imagery or a lengthy
narrative of the conflict, the memorial is elevated above traditional discourse and the typology of the war memorial.

Although the introduction of the angled form came after the original competition process in the design development stage, the rationale for the change was to create an alignment with existing monuments on the Mall. Campbell describes the orientation, "as you descend the path along the wall and reach this angle, you realize that one wing of the black wall points straight at the tall, white Washington Monument a mile or so off, and the other at the Lincoln Memorial, visible through a screen of trees about 600 feet away." Although the overall site strategy within the context of the Mall is effective, the localized effects of the site and monument form provide additional layers of meaning and possibility. While the form of the monument links it to the Washington Monument and the Lincoln Memorial, the traditional ethic of the existing memorials is consciously rejected in many ways. The simple gesture of relating the wall to the existing monuments only in plan signifies once again that only the subtlest historical references are allowed.

Although the monument's form can be interpreted as a response to the fundamental shifts in American society during the Vietnam conflict, the interpretation of cultural change as progenitor of form is effectively undermined by the decision not to present the memorial as a timeline. Again, the historical and cultural contexts, while present, recede to the background intentionally, leaving the wall open to interpretation. The inherent openness validates Pallasmaa's conclusion that architecture must surpass the literal reflection of historical and cultural context by saying, "If we desire architecture to have an emancipating or healing role, instead of reinforcing the erosion of existential meaning, we must reflect on the multitude of secret ways in which architecture is tied to the cultural and mental reality of its time." The cultural and mental reality of the Vietnam Memorial is that of a nation divided. That reality is not challenged, but rather avoided in a series of shrewd decisions regarding the treatment of the surface of the black granite panels.

It must be noted, however, that the shift towards openness is one that is to be experienced at the most intimate level, within arm's reach of the wall. The traditional frontal view from the Mall of the memorial in its entirety still supports the perception of the wall as timeline.

The fact that the wall displays the name of every American soldier killed in the Vietnam conflict is not unusual within the typology of the war memorial, but the method in which the names are presented defies convention and injects meaning into the presentation. Rather than rely on the traditional dictum of alphabetical presentation, the decision was made to present the names chronologically. This simple shift immediately transformed a static display with the dimension of time, giving each name a space in time rather than a place in line. Pallasmaa also recognizes the significance of the representation of time in architecture by stating, "We have a mental need to experience the reality that we are rooted in the continuity of time, and in the man-made world it is the task of architecture to facilitate this experience." So, even though the wall alters the typical perception of chronological presentation, the progression of time is nevertheless present.

The presentation of the names accomplishes what the materiality of the wall denies. While time usually marks buildings with the effects of erosion and weathering, the material choice of black granite, polished smooth attempts to negate those effects. The granite offers permanence, a display that limits the effects of time. Pallasmaa states, "Architecture emancipates us from the embrace of the present and allows us to experience the slow healing flow of time. Buildings and cities are instruments and museums of time. They enable us to see and understand the passing of history and to participate in time cycles that surpass individual life." The decision not to list the names alphabetically also has a humanizing side effect: names are listed in the order they usually used: first, middle initial, last, and without any mention of rank. Names are listed the way a parent speaks a given name, "Michael J Sturgill," rather than the militaristic method of giving the last name first. The names on the wall are presented only as people, not soldiers, leaving open the cause behind their deaths.

While the descending form apparently responds to the amassing casualties and the break in the center reflects the fundamental shift that occurred in 1968, the wall consciously avoids such a literal interpretation and descends in association with funerary ritual. Rather than have the form dictated directly by the number of casualties, the monument unites the beginning and the end of the list in the middle. The literal translation of form becomes the backdrop for an artistic and apolitical interpretation of the events. By uniting the names of the first and the last casualties, the interpretation of the memorial as a simple timeline is subdue, while also allowing simultaneous approach from either end, furthering the idea of openness. Campbell
describes the effect by saying, "the name of the first soldier who died is carved at the angle in the wall, and the names continue to the right in columns in chronological order of date of death, out to the east end where the wall fades into the earth. The names begin again, with the next soldier who died, at the west end, where the wall emerges from the earth. It is as if the wall, after sinking beneath the earth, has continued on around the world underground before emerging once more."  

The experience culminates at the center of the wall, as each wing spreads out to the horizon and the visitor is totally submerged within the memorial. Gaston Bachelard sympathizes with the corner, "To begin with, the corner is a haven that ensures us one of the things we prize most highly – immobility. Consciousness of being at peace in one’s corner produces a sense of immobility, and this, in turn, radiates immobility. An imaginary room rises up around our bodies, which think that they are well hidden when we take refuge in a corner." In an ironical shift from the interpretation of the corner as a cultural breakdown, the wall now offers intimacy and "immobility." Rather than have the wall be a traditional spatial divider, the corner serves to envelop the visitor, holding them in its long arms, while uniting the beginning and the end.

The spatial organization works symbiotically with the tectonics of the wall in that while the sense of intimacy and touch are elevated, as Robert Campbell recalls, "In making this descent you feel you’re entering a cloistered space, set off from the busy surroundings. Streets and skylines disappear to leave you alone with the wall and its names." The visitor is alone with their reflection and the names, and as Paul-Louis Landsberg writes, "For a moment, we have our feet in the land of the dead. A moment later we are once more outside the kingdom of shadows. But during that moment we experienced its bitter cold. And no one is ever quite the same after he has felt it." Implicit is the understanding of each name carrying the weight and size of a person, and with that understanding comes a frightening concept of scale.

The extraordinary power of the names is reflected in the diversity of experience and reactions they generate. Jan Scruggs remembers some of the reactions, "They always touched the names. Fingertips traced out each letter. Lips said a name over and over, and then stretched up to kiss it. Sunlight made it warm to the touch. Perhaps by touching, people regained a sense of life; or perhaps they finally came to peace with death." In evaluating the power of the Vietnam Memorial, there exists a constant tendency to offer statistical analysis: the number of names, the number of panels, the length of each arm, the height of the highest panel, etc. However, all of these facts can be enumerated without the slightest hint of the revealing power of this black granite wall cut into the earth. One must constantly resist the temptation to see as we are taught to see, rather than how our bodies see. Juhani Pallasmaa writes, "The eye is the organ of separation and distance, whereas touch is the sense of nearness, intimacy and affection. The eye controls and investigates, whereas touch approaches and caresses." The memorial is designed both the viewed in its entirety from a typical frontal position, but more importantly, the wall is designed to be experienced intimately, at close proximity, to be touched.
craftsmen and the assiduous hands of its users, seduces the stroking of the hand. The tactile sense connects us with time and tradition: through impressions of touch we shake the hands of countless generations." 18

Visitors are allowed, even encouraged to touch the wall, to connect with the names. The simplicity of the presentation allows a subtle and immediate understanding of the process by which the wall was made. Within this understanding and the subconscious projection undergone with the wall, each visitor bears the responsibility of engraving each name. This simultaneous burden connects the maker, the visitor, and the dead. Pallasmaa states, "As the work interacts with the body of the observer, the experience mirrors the bodily sensations of the maker. Consequently, architecture is communication from the body of the architect directly to the body of the person who encounters the work." 19 This communion is not only an implicit understanding of the process taken to create the inscriptions, but also an empathy and gratitude for the time taken on every letter. The level of craft indicates not only skill, but also reverence.

Even so, the black granite panels provide more than just a medium for the inscription of names. As Robert Campbell recognizes, "At some moment of your visit, probably not at first, you’ve noticed with a slight shock your own face reflected among the names of the dead, an effect that makes the granite mirror a kind of scrim set between past and present, between living and dead, integrating both on a single dark plane." 20 The reflections provided by the wall force the visitor to see themselves at the same time as the names. Jan Scruggs describes the experience, "The panels were like mirrors. The more you looked, the deeper inside you (yourself) you saw. The names floated all around you, along with the clouds." 21 The juxtaposition of the names with the reflections creates an intimacy by association; what is essential is the association of intimacy with the prolonged exposure to one’s own visage.

Pallasmaa articulates the process by saying, "The encounter with any work of art implies a bodily interaction. A work of art functions as another person, with whom one converses." 22 Although the black granite panels offer a distinct reflection of the visitors, it is not so much the reflections that encourage touch, but the names. The incredible power of the name calls from beyond time. In the proper name lies the whole of an individual’s life, and the invocation of their name momentarily brings them back to life, only to have them die again. As Jacques Derrida says, "The name alone makes possible the plurality of deaths." 23 The invocation of the name brings forth the memories of both life and death. Only the name has such simple power to momentarily revive the dead, but each resurrection is bound to the inescapable permanence of death.

Rather than have the name become isolated and fragile, in the wall, the name becomes immortal. Pascale-Anne Broulx, in commenting on the eulogies of Derrida, elucidates the summoning power of the name, "As Derrida has shown in numerous texts, the name is always related to death, to the structural possibility that the one who gives, receives, or bears the name will be absent from it. Mourning thus begins already with the name." 24 The name always carries the implication of death during life, and after death, it brings the memories of that life. The mention or touch of a name is the act of love that Paul-Louis Landsberg recognizes in writing, "A single act of personal love is enough to reveal to me, by creating the presence, or rather the present absence of the person, the essence of human death. One moment of calm in the presence of the dead, even if he be an unknown enemy – and the situation may provoke an act of personal love towards one’s neighbor as such. Then the concept of human mortality recovers its full dignity." 25

Ultimately, the Vietnam Memorial becomes a reflection; a reflection not only of the image of the visitor, but of thoughts, feelings, and beliefs. Visitors are allowed to see themselves and to see inside themselves. Some find grief. Some find triumph. Architecture offers this meditative opportunity by engaging the senses. Connections are made, not only in the psyche, but also in the dermis. As the force of emotional and physical trauma pushes, the wall pushes back, balancing and comforting the visitor. Pallasmaa understands the assuring power of architecture by saying, "Architecture enables us to perceive and understand the dialectics of permanence and change, to settle ourselves in the world, and to place ourselves in the continuum of culture." 26 The Vietnam Memorial provides a home reactions suppressed by traditional memorials; the acceptance of grief, loss, and failure distinguish this wall as a postmodern answer to the discrimination and singularity of modernism.

End Notes:
1 Frascari, Marco. "The Tell-The-Tale Detail."
7 Campbell, Robert. Architectural Record, May 1983 pp 150-151
8 Campbell, Robert. Architectural Record, May 1983 pp 150-151
12 Campbell, Robert. Architectural Record, May 1983 pp 150-151
14 Campbell, Robert. Architectural Record, May 1983 pp 150-151
Throughout the semester, the following questions were posed:

1) Generically, how does cultural dogma affect the production of architecture?

2) What limits, what discriminations, are placed on architecture by status quo convictions?

3) What conditions have been segregated or denied from tectonic manifestation?

4) What is architecture’s role-historically, currently, and prospectively-in accommodating (or refuting) the “unhealthy,” of allowing/disallowing the pursuit of perverse desires?

Jonathan Boelkins, a fourth-year student in the Department of Architecture, was an active participant in this course. Most notably, Mr. Boelkins developed an independent research project that examined issues regarding the architectural and material communication of cultural identity. Mr. Boelkins recently won the School of Architecture Pella student essay competition for this paper. As well, Jon has been named to both the Dean’s List and the Chancellor’s List, is a member of Tau Sigma Delta, received the Wittenberg, Delony, & Davidson Scholarship in 2002, and currently holds a 3.75 overall GPA. As this paper and his academic accomplishments indicate, Jon is a truly accomplished student, designer, and scholar.

Faculty comment:

Mr. Boelkins’ mentor, Assistant Professor Kory Smith made the following remarks about the work:

In the fall semester of 2002, I offered a seminar to upper level architecture students at the University of Arkansas. This seminar, entitled Post-Modern Critiques and the “Unhealthy,” surveyed post-1965 architectural theory and synthesized inquiries concerning the production of architectural ideology. The course facilitated a discussion of diverse (and perverse) social, psychological, and spatial desires of (un)healthiness. This was a questioning of architectural and non-architectural conditions that have been termed “unhealthy.”
THE DIES IRAE ("DAY OF WRATH") AND THETOTENTANZ
("DANCE OF DEATH"): MEDIEVAL THEMES REVISITED IN 19TH
CENTURY MUSIC AND CULTURE

By Erin Brooks
Department of Music

Faculty Mentor: Professor Elizabeth Markham
Department of Music

Abstract:

During the pivotal November 2002 football game of
Arkansas vs. Georgia in the SEC conference championship, the
Georgia marching band struck up their defensive rallying song.
Instead of a typical "defense" song, the band played an excerpt
of the Gregorian Sequence Dies Irae ("Day of Wrath") from the
Roman Catholic Requiem Mass. Drastically dissociated from its
original medieval milieu, this musical Sequence still manages to
elicit the same effect of fear and foreboding nearly a thousand
years later. Precisely because of its deep musical and cultural
roots, the Dies Irae occupies a significant place in history,
closely intertwined from early on with the medieval folk motif
Totentanz ("Dance of Death"), widely depicted in medieval art,
and dramatically revived in 19th century music, art, and literature.

This multi-disciplinary study focuses on the history of art
and music of these two medieval themes during their development,
and then moves on to study them in 19th century culture.
Specifically, the manipulation of the original Gregorian chant
and the incorporation of the idea of a medieval dance are
analyzed in the music of Hector Berlioz, Franz Liszt, and
Camille Saint-Saens. Numerous other contextual links are
explored as well, such as Johann Wolfgang von Goethe, Victor
Hugo, Henri Cazalis, William Blake, and Alfred Rethel, all of
whom created 19th century artistic or literary masterpieces
derived from the thematic seeds of the Dies Irae and the Totentanz.
Although neither of these ideas endured in their original form
during the Romantic era, the inherently compelling nature of
these themes that center on the macabre but inevitable end of life
captivated the Romantic geniuses and continue to intrigue us to
this day.

Editor's note:

Space precluded publication of the entire thesis in this journal. However the work in its entirety can be found on the
Inquiry website.

Chronologically, this study divides into two main areas of
focus; the genesis and permutations of the Requiem Sequence,
Dies Irae, and the "morality dance-with--verses", the Totentanz,
and the revitalization and metamorphosis of these themes in the
Romantic era. Certainly neither the Dies Irae nor the Totentanz
completely ceased to exist in the time between the Middle Ages
and the 19th century. Various alterations in the original ideas
associated with both musical Sequence and the folk motif continually occurred during this lengthy time period. Each generation found its own personal method of dealing with the eternal questions of death and its consequences. Yet the fascination with both the Dies Irae and the Totentanz as products of the medieval mindset has endured for almost a thousand years.

In November 2002, the Arkansas Razorback football team played the Georgia Bulldogs in the Southeastern Conference championship game. At every key defensive point in the game, the Georgia band struck up their rallying song. A typical choice for this kind of "mood" music at a football game is the "Jaws" theme or the "Darth Vader" music - but instead the Georgia band played an excerpted version of the initial section of the Gregorian Sequence Dies Irae ("Day of Wrath") from the Mass for the Dead. Few recognized its original source, but the intent of the music was still clear to all those who heard it. Even today, in 2003, a marching band can play an excerpt from the Dies Irae and generate the same feelings of trepidation and premonition of evil. Twentieth-century composers following the first World War similarly found the Totentanz ("Dance of Death") an effective metaphor for the horror they felt concerning the Nazi regime. Obviously these ancient artistic concepts still hold sway over us today, as an enduring legacy of the human mind's attempts to answer eternal questions. The Dies Irae is definitely no ordinary sacred tune; it carries a rich history of cultural implications, such as the Totentanz, and even achieved an individual importance in music and art throughout the ages.

Initially, the first section of this thesis offers a brief introduction into the history and specific format of the Roman Catholic Mass, and particularly its variant, the Requiem Mass. While the Mass itself took several centuries to coalesce into its precise form, once it achieved this form it became an extremely important cultural institution in Europe. The immense centralized power of the Catholic Church during the medieval era made the Latin Mass an important unifying device across ethnic and linguistic boundaries. The Requiem Mass, ("Mass for the Dead"), was also codified so as to offer the "definitive" Catholic medieval ideology concerning death. As such, the Requiem Mass contained several special components; the Dies Irae was one of these, formally added to the Mass in 1570. This medieval text penned by Thomas of Celano during the late 12th or early 13th century, offers a graphic depiction of the horrors of Judgment Day for sinners. The New Catholic Encyclopedia states that, "The Preface for the Dead" emphasizes the joyful aspects of the Resurrection. The medieval Sequence, however, stresses fear of judgment and condemnation."1

The second chapter of this thesis indeed concerns the background and evolution of the medieval poem, Dies Irae, and also in particular examines the unique Gregorian chant music associated with the text. The origins of the poetic text are traced through both biblical prophecy and the idea of "repentence sermons." The philosophies of redemption and repentance were central to medieval Catholic ideology, and thus the Dies Irae was meant to serve as a warning to both the pious and the wicked. The musical iconography of the Dies Irae indicates this as clearly as the Latin text, if not more so. The rhythm, text-music relationship, modality, method of cadencing, and specific interval patterns all led to complete identification of the text of the Dies Irae with its Gregorian chant counterpart. So much so, in fact, that much later Romantic composers were able to "quote" segments of the music of the Dies Irae to indicate both the text and the general aura of Judgment Day.

The following chapter discusses a related European folk motif, the Totentanz ("Dance of Death"). This medieval movement was widely based in Europe during the Middle Ages; although most prevalent in France and Germany, examples also occurred in Italy and Spain. In its simplest form, the Totentanz was really a dance, albeit one depicted in frescos and woodcuts. The central idea was that death visits us all eventually, regardless of social station or age. Other art forms based on the Totentanz gradually evolved, including elaborate woodcuts depicting every social station, accompanied by texts detailing "Death"'s personal message to each. The Totentanz served as a type of secular yet religiously based motif centered on the uncertainty of life in the Middle Ages. Death was ever-present during these times of plague and war, so communities in the Middle Ages could not afford to shy away from it. Instead, a vital art form was created out of the somehow satisfying idea that death comes to all, regardless of earthly advantages.

The Dies Irae, the Totentanz, and the general belief in Judgment Day were all recreated many times in art during the medieval era. The fourth chapter of this study focuses specifically on the tympanum of Gislebertus at Autun (c. 1130), the Pisan fresco Triunfo della Morte attributed to Bonamico Buffalmauco (c. 1350), Guyot Marchand's woodcuts of a French Dance of Death in 1485, Albrecht Dürer's 1498 Apocalypse woodcuts, Hans Holbein's 1538 set of 41 Dance of Death prints, and Hieronymous Bosch's depiction of Hell in his triptych, The Garden of Earthly Delights, dating from approximately 1510-1515. Each of these medieval artists offered a slightly different outlook on these themes, but all created medieval masterpieces based on the current conceptions of death, Hell, and judgement.

The fifth chapter, "'New Themes' in the Romantic Era," begins the second major section of the thesis. This chapter presents a general overview of the Romantic age, with emphasis on the interconnected nature of the arts during this period. Also of interest are the specific attributes and interests of the Romantics that made revitalization of medieval themes so attractive. Paris in the 1830s is especially important, as many of the artists pertinent to this topic were active there at this time, such as Hector Berlioz and Franz Liszt. Other non-musical artists contributed greatly in both inspiration and encouragement to these composers, so men such as Victor Hugo, Johann Wolfgang von Goethe, and Eugene Delacroix are also discussed. It was the
particular attitude of the Romantic era that allowed for the revival of the *Dies Irae* and the *Totentanz*; these themes dealt with medieval and macabre ideas, which made both of them captivating to many Romantic artists.

Following the Romantic chapter, a group of four successive chapters (Chapters VI-IX) submits most of the musical analysis of the thesis, as well as the most concrete examples of Romantic manipulation of our themes. Respectively, these sections concern Hector Berlioz and the final movement of the Symphonie Fantastique, Franz Liszt and Totentanz: Paraphrase, *ber Dies Irae*, Camille Saint-Saëns' *Danse Macabre*, Franz Schubert's Death and the Maiden, and Gustav Mahler's Symphony No. 2. Berlioz is probably the most famous for his inclusion of part of the chant of the *Dies Irae* in his secular composition. Although this was undoubtedly meant as a sort of parody in view of the program assigned to the symphony, Berlioz’s use of the motive still bears extensive examination. It is extremely interesting in its rhythmic, thematic, and modal characteristics. Berlioz does not merely quote the chant verbatim; he instead rhythmically augments it, creating a similar feeling to the augmented bass line of a motet. The same “lengthening” phenomenon has been documented in other types of Western music as well as in Eastern music, such as medieval Chinese dance-tunes performed at the Japanese court today. The Symphonie Fantastique also creates an ingenious mixture of perceptions, as it fuses the fragment of the *Dies Irae* with a wild witches’ round dance—perhaps an indication of the universality of “dance” as a part of death.

Franz Liszt’s Totentanz is quite different; this composition is an incredible melding of a theme and variations and a virtuoso piano concerto. Liszt was very much influenced by Berlioz’s symphony, yet Liszt’s devout Catholicism likewise led him in new directions involving the *Dies Irae*. The first three lines of the Gregorian chant serve as the “theme” in Liszt’s work, which are then split off and treated in a wholly Romantic fashion. Liszt also composes the same type of rhythmic augmentation as Berlioz, although not as extreme. Furthermore, the somewhat modal character of the beginning of this piece at least suggests that Liszt attempted to be historically accurate in his initial statements of the Gregorian sequence.

Camille Saint-Saëns was active slightly later as a composer than Hector Berlioz, and was a younger associate of Franz Liszt’s. However, his music bears the influence of both in many ways. His symphonic poem, *Danse Macabre*, plays upon several threads of previously mentioned themes. He uses the Totentanz of Goethe and the *Danse Macabre* by Henri Casalis as inspiration—both of these 19th-century poems exhibit the relatively new Romantic idea of skeletons dancing in the graveyard at midnight. But Saint-Saëns too includes a musical quote of the *Dies Irae* in his *Danse Macabre*. He also manipulates the interesting ethnomicological association of Death as a violinist or fiddler. This motif is common in many cultures, and is briefly discussed during this chapter.

The final musical analysis section features short commentaries on the works of two other 19th-century composers, Franz Schubert and Gustav Mahler. Neither uses the *Dies Irae* or the Totentanz as straightforwardly as the previous three composers. Yet the works of these two men submit interesting adjuncts to the main topic of this study. Franz Schubert’s Death and the Maiden, a short 19th-century lied, proffers an alternative to the view of “Death” as a terrifying experience. This song instead depicts a young maiden who is seduced by the calmness and certainty of “Death”—he will take her to a kinder place than this cruel world. The idea of death as alleviating earthly pain was present even in the Middle Ages, and Schubert’s song reinforces this surrogate view of our ultimate end. Gustav Mahler’s Symphony No. 2 is a much more complex and personal use of the *Dies Irae* in music. This massive symphony creates a programmatic journey through death, remembrance, fear, resurrection, judgment, and finally, a move to eternal bliss. The *Dies Irae* is used as a mighty symbol of the day of reckoning that Mahler felt all had to face before the ultimate salvation. This highly personal interpretation of the “Day of Wrath,” reflects Mahler’s individual ideology as well as a late Romantic example of gargantuan symphonic music.

The final chapter of the body of this work affords several examples of how the *Dies Irae* and the Totentanz were promulgated in non-musical 19th-century art. Specifically, the artwork of William Blake, Victor Hugo’s poem *The Dance of Demons*, Goethe’s poem Totentanz, and Alfred Rethel’s woodcuts of Death as a Cutthroat and Totentanz Auch ein Jahr 1848 are considered in this section. The Romantic era seemed to particularly value creativity and individuality, so it is not surprising that all of these artworks are conspicuously characterized by their creator. Yet, the common threads running through these works of art as well as the earlier mentioned musical compositions are remarkable; the incredible renewal and concentration on the *Dies Irae* and the Totentanz during the Romantic era is unique in European cultural history.

The two themes of the *Dies Irae* Sequence and the Totentanz have both enjoyed a long, intertwined history. Each evolved as different reactions to the universal human questions of what happens after death. However, the longevity and recycling of these ideas throughout the history of art of all kinds is a testimony to their artistic worth. The *Dies Irae* and the Totentanz were not constantly quoted and used solely because they dealt with death; both are valid artistic triumphs of the human intellect. The *Dies Irae* survived not only because of its subject matter, but also because it is a beautiful, symmetric poem set to an equally appropriate and balanced melody. The Totentanz spread over Europe as a popular secular movement because of its intrinsic message, but also offered a wonderful glimpse into the complicated social hierarchy of the late Middle Ages. The later, concentrated use of both of these originally medieval themes in 19th-century orchestral masterpieces reflected the fascinating
social mores and ideals of the Romantic mindset. The *Dies Irae* and the *Totentanz* offered the works of Berlioz, Liszt, and Saint-Saëns both a grounding in a rich artistic tradition and an aid to “Romanticize” the eternal themes of life and death.

End note:


Faculty comment:

Ms. Brooks’ faculty mentor, in her letter to the publication board extolling the importance of Ms. Brooks’ research, makes the following comments:

Erin’s chosen topic for her Honors Thesis in Music History - the borrowing of the medieval chant *Dies Irae* (“Day of Wrath) from Mass for the Dead and the interlacing of the medieval *Totentanz* (“Dance of Death”) as *topoi* in Romantic orchestral works - has led her deep into interdisciplinary, cross-cultural study. Tracing the *Dies Irae* and the *Totentanz* through the visual, literary and performing arts of medieval Europe - the concern of the first part of her thesis - involved reading widely beyond the field of traditional musicology (in biblical and liturgical studies, in art history, folklore, literary studies, historical dance studies, historical ethnomusicology, and musical iconography). In respect to the *Totentanz* in particular, this also required extensive dealing with the mainstream scholarly studies in German, Dutch, and French. Her final account of the medieval development of these themes is a well researched and skillful, selecting and drawing together from a vast (and in certain aspects controversial) scholarly field of those strands relevant for her particular emphasis/evidence, and especially on iconographical and structural musical grounds, for why an ancient chant associated with death rituals of the Roman Catholic Church can endure so tenaciously, inspire artists across centuries, and still evoke in us today the emotions it was presumably intended to evoke at the time it came into being.

The second part of the thesis is devoted to the remarkable surge of interest in both the *Dies Irae* and the *Totentanz* among leading artists, composers, and writers of the Romantic Period. Just as the inextricability of the two themes in the medieval period demands a multi-pronged approach to their study, the unified artistic movement of the Romantics in general, and intellectual cross-fertilization ensuing from contacts and friendships among the key figures in this story in particular, also require broad-based enquiry. Erin again draws from across many fields to address both themes as borrowed in 19th century music, art, and literature. While the letters of the three principal composers who worked with the *Dies Irae* and/or the, *Totentanz*—Berlioz, Liszt and Saint-Saëns—are central to her fascinating accounts of how each came to borrow the medieval themes, of direct and indirect mutual influencing, and of contacts with and inspiration from Goethe (and especially from *Faust*), Hugo and others, her technical musical analyses of relevant passages from Berlioz’ *Symphonie Fantastique*, Liszt’s *Totentanz* (*Danse Macabre*), and Saint-Saëns’ *Danse Macabre* provide explicit demonstrations of how the modal melody of the *Dies Irae* is actually used in Romantic tonal environments. Erin makes a number of original and thought-provoking observations. Perhaps among the most significant of these is that the perfect symbiosis of music and text in the original chant is so powerful that even in “wordless” 19th century idioms, text-structure holds sway. Consequently the overwhelmingly strong association between “death, darkness, and the underworld” and the key of D-minor may well be linked ultimately with the mode of the medieval Requiem chant, *Dies Irae*.

While the historical ethnomusicologist in me probably lured Erin far afield from her intended path, particularly in her following the *Totentanz* and its relationship with actual European dance-types (folk and courtly) of the time, and (certainly by the end of her research) in her fearless taking-on of secondary literature in German, French, and even Dutch, what has come out in the multidisciplinary wash is a very fine and cohesive study. Indeed, a good deal of what is covered as contextualization has not been dealt with in such detail in English before. This alone indicates the value of the essay. But Erin’s particular approach to the way text/melody balance intrinsic to the original *Dies Irae* chant perseveres at the structural level when “borrowed” and incorporated by Romantic composers will probably stand as the hallmark of her original contribution to this piece of musical history. Here it is the underlying structure that defies change and, no matter in which other genres and contexts it may appear and whatever encrustations it may now bear, testifies to the “old” and the “new” being “the same piece”. I take Erin’s lead in pointing to the history of the *Dies Irae* in our own musical tradition as an instance of an almost 1,000 year-old liturgical item that has persevered as a musical structure both in the Requiem Mass and in the secular domain - admittedly assisted by its particular poignancy and power for the ephemeral human - to be significant. While the undergraduate textbook in Music History may simply acknowledge that Berlioz, Liszt, and Saint-Saëns “quote” from the Dies Irae, Erin has delved far beneath the surface, even into the psychology of the composers themselves, to try to elucidate how and why this particular chant evidently holds such a spell over us.
THE DIES IRAE (“DAY OF WRATH”) AND THE TOTENZAN (“DANCE OF DEATH”): MEDIEVAL THEMES REVISITED IN 19TH CENTURY MUSIC AND CULTURE

By Erin Brooks
Department of Music

Faculty Mentor: Elizabeth Markham
Department of Music

Abstract:

During the pivotal November 2002 football game of Arkansas vs. Georgia in the SEC conference championship, the Georgia marching band struck up their defensive rallying song. Instead of a typical “defense” song, the band played an excerpt of the Gregorian Sequence Dies Irae (“Day of Wrath”) from the Roman Catholic Requiem Mass. Drastically dissociated from its original medieval milieu, this musical Sequence still manages to elicit the same effect of fear and foreboding nearly a thousand years later. Precisely because of its deep musical and cultural roots, the Dies Irae occupies a significant place in history, closely intertwined from early on with the medieval folk motif Totentanz (“Dance of Death”), widely depicted in medieval art, and dramatically revived in 19th century music, art, and literature.

This multi-disciplinary study focuses on the history of art and music of these two medieval themes during their development, and then moves on to study them in 19th century culture. Specifically, the manipulation of the original Gregorian chant and the incorporation of the idea of a medieval dance are analyzed in the music of Hector Berlioz, Franz Liszt, and Camille Saint-Saëns. Numerous other contextual links are explored as well, such as Johann Wolfgang von Goethe, Victor Hugo, Henri Cazalis, William Blake, and Alfred Rethel, all of whom created 19th century artistic or literary masterpieces derived from the thematic seeds of the Dies Irae and the Totentanz. Although neither of these ideas endured in their original form during the Romantic era, the inherently compelling nature of these themes that center on the macabre but inevitable end of life captivated the Romantic geniuses and continue to intrigue us to this day.

Prologue

December 7, 2002- The pivotal football game of the Arkansas Razorbacks versus the Georgia Bulldogs in the Southeastern Conference Championship is underway.

Unfortunately, the underdog Razorbacks are immediately decimated by Georgia’s potent defense at the beginning of the game. Surrounded by 75, 835 screaming fans, the
University of Georgia Redcoat Marching Band attempts to fire up the crowd even more. After every defensive stand by the Georgia football team, the marching band strikes up their defensive rallying cheer. A typical choice for this kind of “mood music” is the “Jaws” theme or the “Darth Vader” theme from Star Wars, two common defensive songs. Instead, the University of Georgia marching band plays an excerpt from the Gregorian chant Sequence Dies Irae (“Day of Wrath”) from the Roman Catholic Requiem Mass. Certainly not every one of the fans present recognizes the derivation of this tune. A few persons may place it as part of Hector Berlioz’s borrowed rendering of the Dies Irae in his Symphonie Fantastique; to other fans it may merely sound slightly familiar. Yet the enormous crowd immediately perceives the intended emotional effect of this music. But how is it that a Gregorian chant so far removed from its original context still has meaning? Significantly, this particular Sequence holds a singular place in the history of Gregorian chant. The unique aural qualities of even a short excerpt of the Dies Irae provide an evocative example of how perfectly this music portrays the text that accompanies it. Even in drastically divorced settings, the music of the Dies Irae is able to indicate the mood of its medieval text. Evidently, the Dies Irae is no ordinary sacred tune; it carries a rich intertwined history and even achieved an individual importance in secular music, art, and literature. While not readily recognized in current scholarship in the separate disciplines of art history, literary and liturgical studies, etc., from a music-based approach, it appears that the Dies Irae was intricately intertwined with another medieval theme, the Totentanz (“Dance of Death”). This secular folk motif, associated with actual dance-forms of the time and manifest in both the literary and visual arts, was also obsessed with Death and the eternal price for sin. Indeed, the intense interest in both
the Dies Irae and the Totentanz exhibited by the Romantics— including writers and artists, but especially composers— provided the initial rationale for tracing the two themes from the point of their being merely different representations of a single idea.

Chronologically, this study divides into two main areas of focus; the genesis and permutations of the Requiem Sequence, Dies Irae, and the “morality dance-with-verses”, the Totentanz, and the revitalization and metamorphosis of these themes in the Romantic era. Certainly neither the Dies Irae nor the Totentanz completely ceased to exist in the time between the Middle Ages and the 19th century. Various alterations in the original ideas associated with both musical Sequence and the folk motif continually occurred during this lengthy time period, as each generation found its own personal method of dealing with the eternal questions of death and its consequences. Yet the fascination with both the Dies Irae and the Totentanz as products of the medieval mindset has endured for almost a thousand years. A marching band can play an excerpt from the Dies Irae (therefore from the Roman Catholic Requiem Mass) and generate the same feelings of trepidation and premonition of evil. Twentieth-century composers following the first World War similarly found the Totentanz (“Dance of Death”) an effective metaphor for the horror they felt concerning this apocalyptic struggle. Obviously these ancient artistic concepts still hold sway over us today, as an enduring legacy of the human mind’s attempts to answer eternal questions.

Part I: Intertwined Themes of the Medieval Age

While the Mass itself took several centuries to coalesce into its precise form, once it achieved this form it became an extremely important cultural institution in Europe.

The immense centralized power of the Catholic Church during the medieval era made the
Latin Mass an important unifying device across ethnic and linguistic boundaries. The Requiem Mass, ("Mass for the Dead"), was also codified so as to offer the "definitive" Catholic medieval ideology concerning death. As such, the Requiem Mass contained several special components; the *Dies Irae* was one of these, formally added to the Mass in 1570, but known as a musical item in the earlier part of an extant collection of notations from the 12th, 13th, and 15th centuries, and probably having even earlier musical roots. The medieval text itself, possibly penned by Thomas of Celano during the late 11th or early 12th century, offers a graphic depiction of the horrors of Judgment Day for sinners. The New Catholic Encyclopedia states that, “The Preface for the Dead…emphasized the joyful aspects of the Resurrection. The medieval Sequence, however, stresses fear of judgment and condemnation.”

References in the poetic text of the Sequence may be traced through biblical prophecy, especially in the prophecies of Zephaniah and Isaiah, both strongly linked with the idea of a wrathful God, as well as the idea of “repentance sermons.” Still, the rewards for those who lived a just life were also evident by this point in time. The philosophies of redemption and repentance were central to medieval Catholic ideology, and thus the *Dies Irae* was meant to serve as a warning to both the pious and the wicked. The musical iconography of the *Dies Irae* indicates this as clearly, if not more so, than the Latin text. The distinctive melody to which the *Dies Irae* is set became so recognizable that hundreds of years later, 19th century composers were able to quote parts of it to create a

---

1 Kees Vellekoop, *Dies Ire, Dies Illa: Studien zur Frühgeschichte einer Sequenz* Bilthoven: A.B. Creyhton, 1978. pp. 62-76, 185. Vellekoop codified his version of the *Dies Irae* from the following manuscript: Neapel, Biblioteca Nazionale- Hs. VII D 36, f. 16r. Origin: Carmaninico, Benedictine. Dates from the end of the 12th century (oldest part). In the 15th century manuscripts from the 12th, 13th, and 15th centuries were bound together in a collection- this *Dies Irae* dates from the 12th century and is contained in this collection. (See Vellekoop, p. 29).
specific effect in their compositions. The *Liber Usualis*, a compendium of prayers, lessons, and chants for the more important services of the Catholic Church, includes the entire text and melody of the *Dies Irae*. In addition, Kees Vellekoop has codified the standard version of the Sequence melody from the earliest surviving notated sources (Figure 1). There are several unique musical characteristics of the Sequence melody of the *Dies Irae*. The most distinctive segment of the melody is definitely that carrying the first two line of the Sequence, with the majority of later composers choosing to quote these lines nearly verbatim from the original chant. Rhythm, text-music relationships, modality, method of cadencing, and specific interval patterns all lead to complete identification of the text with its musical counterpart. Vellekoop argues that one of the reasons for the success and longevity of this Sequence is the perfect symbiosis between the music and the text. Vellekoop includes a lengthy diagram matching the stresses of the syllables with their counterparts in the chant melody and also provides a measured reconstruction of the way the rhythm might be performed (Figure 2). Of particular importance is the fact that each of the text-lines can be divided into two roughly four-"beat" sub-segments. The first line offers a good example: “Di-es ir-ae, Di-es il-la” is paired melodically with single quarter notes. Thus the melody naturally falls into a declamatory pattern that seems to imply subdivisions of both four and eight-beats. One specific reason for this is the unusually syllabic nature of the *Dies Irae*. Gregorian chant has traditionally been thought to be somewhat rhythmically amorphous, if for no better reason than the fact that the rhythm is not notated, or that if embedded has not been fully recognized by chant scholars. Early medieval performance practice is a tricky field, so

---

we really still have an incomplete idea of how rhythmic or arrhythmic chant really was. Yet, this Sequence has a more inherently rhythmic structure than practically any other extant example of Gregorian chant. *Dies Irae* tends to avoid long melismas (with one word or syllable stretched out over many notes), and instead focuses on a simple, almost declamatory style of chant. This was possibly done intentionally by the original composers of the Sequence, in order to increase both its solemnity and dramatic effect: the important text would be more clearly articulated syllabically and neumatically than if set melismatically. Certainly the skillful matching of the text and the melody is one attribute that must have helped the *Dies Irae* survive the Catholic Church’s great purge of sequences from its Masses during the Council of Trent in the 15th century. Only four sequences remained liturgically permissible after this council; the important subject matter of the *Dies Irae* Sequence probably also helped its chances.

Finally, the *Dies Irae* simply offers a suggestive, easily recognizable melody. Perhaps the fact that the music is so evocative of the text is part of the reason why 19th century composers were drawn to this chant. Thus the singular structure and aural qualities of the *Dies Irae* offered later composers of “wordless” orchestral works an unique opportunity to suggest textual implications through music alone.

The origins of a “death dance” are a fascinating study in cross-cultural exchange, as this theme has echoed through various European cultures at different times in history. The *Totentanz* movement was widely spread during the Middle Ages; although most prevalent in France and Germany, examples also occurred in Italy and Spain. While literary seeds for the *Totentanz* have been suggested, in its simplest visual forms as murals and frescoes in graveyards and churches, the *Totentanz* was really depicted as a dance towards death, or towards the Charnel House (or *Beinhaus*, “Bone House”) in the medieval cemetery (Figure 3).
In the earlier Totentänzen, the procession towards death was depicted as a reigen, a dance category that includes both round-dances and chain-dances. The frescos in the church of Kermaria-in –Isquit dating from 1430 feature this kind of Totentanz (Figure 4). Literally a “reigen of death”, this fresco features couples, in each of which a representative of a trade, profession, or social class holds hands with a depiction of “Death,” who is consistently pictured as a skeleton. This kind of stately reigen, holding hands and
dancing towards the ultimate destination of death, contrasts greatly with the more frenzied kind of death dance found in some other Totentänzen. Thus there were two sides to the Totentanz: the procession or escort into death, and the wild, grotesque horror of the people surprised and dragged into death. This duality also fits in with the idea that sometimes Death comes as a friend and sometimes as a horrible enemy.

Through later Totentänze the chain-dance with held hands gradually moved to a procession of pairs, by which point the earlier reigen had all but disappeared from the visual aspects of the Totentanz. These developmental changes in the theme corresponded with those in actual dances of the era; from reigen, and in particular the branle, to basse danse and on to pavanne. The central idea was that death visits us all eventually, regardless of social station or age. Other art forms based on the Totentanz gradually evolved, including elaborate woodcuts depicting every social station, accompanied by texts detailing “Death”’s personal message to each. The Totentanz served as a type of secular yet religiously based motif centered on the uncertainty of life in the Middle Ages. Death was ever-present during these times of plague and war, so communities in the Middle Ages could not afford to shy away from it. As reasons for the rise of the Totentanz Reinhold Hammerstein offers widespread medieval ideas such as death as punishment for sin, fear of an unprepared death that would lead to Hell, the devil’s escorting of the damned to Hell in Judgment Day iconography, and the associations of the Devil, Death, and the powers of the Musician. The evolution of the visual, thematic, musical, and literary topos of the Totentanz may thus be seen as an intriguing adjunct to

---

the idea of the “Day of Wrath.” In particular, the links between Judgment Day iconography and music serve to connect the Dies Irae and the Totentanz.

Even in the earliest representations of the Totentanz, Death was often portrayed as playing a sort of straight-trumpet; Hammerstein suggests this is because of traditional Judgment Day iconography of the “trumpet sounding” to call souls to be judged.\(^4\) However, Death was also often pictured playing “low”, dance instruments such as the kettledrum and various pipe and reed instruments (Figure 5).

![Figure 5. Reinhold Hammerstein. Ibid. Plate 231. “Bleibach, Beinhauskapelle, Totentanz, 1723: Todesmusikanten.”](image)

Thus, the visual representations of the Totentanz present a dichotomy between the actual “dance of death” and the more stylized association with Judgment Day and the Dies Irae. The verses that usually accompanied Totentänzen reflected this connection as well; Hans Holbein the Younger’s set of 41 woodcuts offers a particularly good example of this type of textural collage. His woodcuts feature a Danse Macabre, but also include a “Last Judgment”, under which the verse states:

Before the mighty Judge’s chair Comes
reckoning for each man alive; Fear then,
the judgments rendered there: You know
not when He will arrive.  

Holbein is in fact only one of many artists during the medieval era that utilized these two medieval themes. The strength of the ideology concerning the “immortal soul” was immense during the medieval era, and the Catholic Church at this time consolidated its power by using images incorporating these themes to influence the largely illiterate population of Europe. Thus, there are a wealth of pictorial examples of the “Day of Wrath”, the Totentanz, and Judgment Day dating from the medieval era. The Dies Irae, as an actual part of the Roman Catholic liturgy, was more of an “official” response to the eschatological question, whereas the Totentanz operated mainly as a folk motif, albeit one patronized by the Catholic Church. Famous medieval masterpieces such as Gislebertus’s tympanum portal at Autun, Bonamico Buffalmacco’s Triunfo della Morte in the Campo de Santo at Pisa, Guyot Marchand’s 1485 Dance of Death woodcuts, Albrecht Dürer’s 1498 Apocalypse woodcuts, and Hieronymus Bosch’s famous triptych, The Garden of Earthly Delights (ca. 1510-1515) depict various permutations of these ideas.

**Part II: The “New” Themes of the Romantic Age**

More than six centuries later, the Romantic age rediscovered the “romance” of the Middle Ages. For some musicians, artists, and writers, these “new” themes offered an exciting, innovative way to create highly personal art. Probably the most important aspects of Romanticism, however, were the synthesis of new and old ideas and the
merging of all the arts. For the first time, painters, playwrights, composers, poets, and actors were all part of a great movement that was independent of the old “hypocrisy” engendered by the aristocracy. It was this world, of men (and women) such as Hector Berlioz, Victor Hugo, Franz Liszt, Johann Wolfgang von Goethe, Eugene Delacroix, and many others, that became fascinated with the ideals of the ages of the distant past. A remarkable musical consequence of this interconnected movement was the new focus on music pertaining to both ancient and modern literature. The idea of expressing words or stories through music was of course an age-old rationale for the composition of music. This type of musical association continued during the Romantic era as well, but new genres of music linked with literature arose during the 19th century, such as program music, the symphonic poem, and art songs. In particular, the inspiration for the symphonic poems and program music is a telling portrait of the Romantic mindset. The unique atmosphere of the Romantic era, then, is what led to the revival of medieval themes such as the Dies Irae and the Totentanz: these themes dealt with medieval and macabre ideas, which made both of them captivating to many Romantic artists. Specifically, in three non-liturgical musical works by Hector Berlioz, Franz Liszt, and Camille Saint-Säens, both the Dies Irae and the Totentanz become reinterpreted though the unique Romantic perspective.

The music and ideas of Hector Berlioz shaped the mid-to-late 19th century attitudes toward art, influencing both his generation and those who followed him. The most famous example of Berlioz’s genius for orchestral color is the Symphonie Fantastique, an early work written in 1830. While this entire piece is a showcase for Berlioz’s compositional innovations, only the fifth movement, Songe d’une nuit du

---

"Sabbat" (“Dream of a Witches’ Sabbath”), uses the Sequence *Dies Irae*. The connection with the Witches Sabbath introduces the Walpurgis Nacht as yet another motif of evident allure for the Romantics. Berlioz’s reasons for including this unorthodox version of the *Dies Irae* appear to be manifold; it is frequently explained as a grotesque and symbolic parody of a well-known religious theme meant to shock the audience.

Another rationale for Berlioz’s use of the *Dies Irae* in his secular composition was the increasing Romantic interest in mythical and fantastic ideas. Although this has been over-generalized in works about the Romantic era, many artists, writers, and musicians were suddenly interested in creating works about the Middle Ages, fantastical ideas, and myths. Jacques Barzun associates several of these threads of Romanticism with Berlioz’s composition of the *Symphonie Fantastique*. He wrote:

Specifically, the transmogrified appearance of the beloved in the finale of the *Symphonie Fantastique* is a precipitate of several suggestions: in Goethe’s *Faust*, the Brocken scene includes a ghostly appearance of the heroine, and later when she recognizes her guilt in church, the *Dies Irae* sounds, as in Berlioz’s symphony. Not long before Berlioz began his piece, Hugo had published a volume of verse in which a *Ronde du Sabbat* is described in detail with its specters, beasts, and witches reveling in a mockery of religion: in Berlioz the *Dies Irae* is parodied. Finally, in the early months of 1830, Berlioz may have read De Quincey’s *Opium Eater* in Musset’s translation and adapted one of its ideas.¹

Barzun’s mention of Goethe’s *Faust* is especially meaningful for both medieval themes (*Dies Irae* and the *Totentanz*). Not only did the *Dies Irae* occur in the Cathedral Scene of
Faust before Walpurgis Night, a festival traditionally allied with “witches’ orgies”\(^7\), but Goethe also penned the poem Totentanz. The influence of Goethe upon both Berlioz and Liszt cannot be overstated. The friendship between these two composers was actually in part sparked by Goethe’s writings. Berlioz wrote, “On the day before the concert, Liszt called on me. It was our first meeting. I spoke of Goethe’s Faust, which he confessed he had not read, but which he soon came to love as much as I. We felt an immediate affinity, and from that moment our friendship has grown ever closer and stronger.”\(^8\)

The fact that Berlioz used the Dies Irae in this movement is widely recognized, yet he did not simply “quote” from the original Gregorian chant version. In the annotations within his score Berlioz was quite specific as to where the Dies Irae and the Witches’ Round Dance occur (and when the two are combined), and the manner in which Berlioz accomplished this musical quotation, fugue, and combination is fascinating.

Initially Berlioz chose not to quote the entire 19 strophes of the medieval sequence; instead he mostly used variations on the opening two “phrases,” that carry the first two lines of the first strophe. Respectively, these follow the text-lines:

\[
\begin{align*}
\textit{Dies irae} \ & \textit{dies illa} \\
\textit{Solvet seclum in favilla}
\end{align*}
\]

Kees Vellekoop provides a measured reconstruction of the Dies Irae strophes into modern Western staff notation (see earlier reference). Each text-line in the model strophe is broken into quarter-note beats (or their equivalents), each carrying a single syllable of the 8-syllable text-line. Although the articulations of some syllables include what we
would consider “neighbor” notes or other small embellishments, the basic structure in this reconstruction contains 8 beats per phrase/ text-line.

The equivalent of the first two text-lines of the Dies Irae are initially stated in dotted half-notes in bassoons and ophecleides. This segment of the chant is then reiterated in mm. 147-157, except in dotted quarter-notes, halving the durational values of the first statement. In addition, this reduction in dotted quarter-notes starts on beat 2 of m. 147, creating a slightly lopsided feeling in the 6/8 meter. A further proportional reduction into the “6/8 dance-meter” associated with the Witches’ Round- Dance follows as a brief third illustration of the basic Dies Irae motif in mm. 157-161 (Figure 6). The upper strings and woodwinds exhibit a harmonized version of the first two lines of the chant, again, but the interval between the first two notes is generally a major 2nd, instead of the minor 2nd typically present in the initial segment of the chant. Also, the rhythm is further diminished in this third statement, as the text-lines are played in eighth and quarter notes. This rhythm completely alters the original equalization approach in the first two statements of the Dies Irae theme; however, the misplaced accents are retained because this third example again does not start on the strong beat of the measure.
Here, Berlioz’s rhythmic treatment of the Dies Irae quote is interesting in the manner in which it is systematically “contracted” or diminished in meter. The intervals are what makes this quote hold together, as the rhythmic manipulation is somewhat difficult to hear. This contraction is further obscured by the fact that each of the diminutions occur on a weak beat, displacing the accents in both the meter and the “expected” accents of the Dies Irae quote. Each of the three main statements derived from the Gregorian chant is treated in this same rhythmic manner. Berlioz’s application of a version of the renowned
Dies Irae Sequence in his secular symphony was undoubtedly a novel idea, yet the real genius of this movement lies in the way he was able to combine this motive with the Witches’ Round-Dance in such a cohesive manner. Structurally, the Dies Irae precedes the Ronde du Sabbat, which follows in a fugal treatment, then finally the two are combined (Dies Irae et Ronde du Sabbat ensemble). Moreover, each of these two themes is “framed” and foreshadowed by the other at every major structural junction. Thus, the Witches’ Dance is “foreshadowed” before the entrance of the Dies Irae, the three entrances of the Dies Irae are divided by chromatic runs reminiscent of the Witches’ Dance theme, and the Witches’ Dance is also interrupted by the Dies Irae. The orchestration and the treatment of both themes in this final section of the Symphonie Fantastique is significant as well. The Dies Irae is here mostly confined to the woodwinds and brass, while the strings are always associated with the witches’ dance. This division seems important, if merely for the fact that the brasses sound more like religious music than does the “profane” string section. Here again is a possible link between the diabolical connections of string instruments as opposed to the signaling capacity of the brass (especially trumpets).

Hector Berlioz revolutionized both symphonic music and the idea of “programme music” with this immense symphony. Even today this symphony is regarded as one of the early masterpieces of Romanticism, and is still startling in its use of the orchestra. The genius of this work definitely rests partially on the idee fixe and its program, but the unique brilliance of the final movement, with its complicated intertwining of the Dies Irae and the Ronde du Sabbat, has also undoubtedly led to the Symphonie Fantastique’s enduring success.
This famous symphony fascinated Franz Liszt, the legendary piano virtuoso and composer, who transcribed it for piano and spread its fame over Europe. Liszt was initially heavily influenced by Berlioz’s skills in orchestration, yet his attraction to the Dies Irae is even more multi-layered than Berlioz’s. Liszt was drawn to the Dies Irae because of his own fervent Catholicism, his captivation with the Faustian legend, Johann Wolfgang von Goethe’s poem Totentanz, and the powerful affect of the Triunfo della Morte fresco in Pisa. This complex framework of thematic contributions led to the composition of his Totentanz: Paraphrase über „Dies Irae,” a gargantuan work for solo piano and orchestra. This work is an unique example of Liszt’s compositional style, a piece that seems strangely torn between its theme and dramatic displays of pianistic virtuosity. Other important analytical concepts include how the Dies Irae plainchant is exploited as motivic tool in Totentanz, specific important treatments in the theme, Variation IV and the fugato, as well as Liszt’s tonal characterization of the Dies Irae.

In his use of the Sequence, Liszt followed the original notation of the plainchant more closely than did Berlioz. While Berlioz abandoned the apparent perfect symmetry of the original 8-beat phrases in the Symphonie Fantastique, Liszt retained the natural 4-beat divisions dictated by the text-structure of each text-line of the Sequence. Liszt also appears to have only used the first strophe of the Dies Irae in his Totentanz. He split these three phrases into their 6 respective 4-beat sections. The same 4-note cadential formula is present in “in favilla” and “cum sybilla,” so Liszt was really only left with five separate motivic fragments (Figure 7). The five motivic fragments are worked out rhythmically, harmonically, and intervallically during the course of Totentanz. Each 4-beat motive is easily aurally recognizable, so that Liszt was able to use these motives as
the linking entity for this massive piece. The formal structure of Totentanz is more complex, as Liszt initially commits to a theme and variations structure, but later “breaks down” this idea into increasingly complicated sets of variations and cadenzas. Thus, while the chosen formal structure does lend some musical integrity to Totentanz, it is really the motivic fragments of the Dies Irae that serve to unify this piece. Totentanz is a brilliant work of art in its own right, but the complex web of thematic inspirations for this composition is equally enthralling. Liszt’s work was perhaps the culmination of the intermingled Romantic revisitations of the Dies Irae and the Totentanz. Liszt himself was the perfect composer to realize this intermingling. As a deeply religious man, yet attracted by the macabre aspects of life, Liszt understood the warnings that the original ideas of the Dies Irae and the Totentanz contained for both the pious and the wicked.

Figure 7. Motivic Fragments in Totentanz

Although Camille Saint-Saëns had his own unique compositional style, he was heavily indebted to earlier Romantic composers such as Berlioz and Liszt in both his oeuvre and inspiration. Saint-Saëns first met Liszt in 1851, when he was only fifteen years old. The two would later cultivate an immense professional admiration and friendship until Liszt’s death. Likewise, Berlioz and the young Saint-Saëns were acquainted in the Parisian musical scene. These two composers were perhaps
instrumental in Saint-Saëns’ interest in the *Dies Irae* and the *Totentanz* (*Danse Macabre*), but he was inspired by other sources as well. Saint-Saëns originally conceived a song based upon the poem of Henri Cazalis entitled “*Danse Macabre, or Egalité, Fraternité.*” This poem featured the midnight revel of skeletons dancing in the cemetery, a purely 19th century idea that originated with Goethe’s poem, *Totentanz*. However, Cazalis’s concept of death as the great leveler still retained some thematic aspects of the medieval *Totentanz*. Saint-Saëns’ initial song was reworked in 1874 into an elaborately orchestrated symphonic poem, *Danse Macabre*. This new *Danse Macabre* definitely exhibits Lisztian influence, but it is uniquely Saint-Saëns’ in the inspiration to combine so many different artistic threads. *Danse Macabre* integrates the ideas of Death as a fiddler, the medieval *Totentanz*, Goethe’s *Totentanz*, and the *Dies Irae*.

The idea of Death as a fiddler, or more generally as a musician, is an extremely old and complex motif. Early *Totentanz* frescos often depicted Death as a musician, with the power to “call” people to him with his playing. The idea of the violin in particular as a diabolical instrument is a much more complex issue. In many European beliefs, Death (or a demonic individual) is the master fiddler who can enchant you with his playing. Saint-Saëns’ casting of Death as the fiddler in *Danse Macabre* draws upon this long association. Saint-Saëns’ symphonic poem was also allied with the medieval *Totentanz* through its use of a round-dance rhythm, and with Goethe’s *Totentanz* as a result of the midnight dance of the skeletons, which Saint-Saëns imaginatively portrayed in his novel orchestral effects.

The inclusion of the *Dies Irae* in *Danse Macabre* creates a much more complex piece, both thematically and musically. The *Dies Irae* becomes one of the three themes
of this work, which are respectively, the chromatic solo violin theme, the dance theme, and the *Dies Irae*. The Gregorian chant is altered here in several significant ways. The two initial 4-note motives are spelled in way that creates a drastic change from the original intervallic pattern; while the Hyperdorian note-set of the original contained a minor 2\(^{nd}\) between the first two notes of motive 1 and a major 3\(^{rd}\) between the first two notes of motive 2, now certain notes are altered to yield a major 2\(^{nd}\) and a minor third, respectively. Likewise, the rhythm is changed, creating a lopsided feel due to the accents found on the “wrong” beats in the overarching triple meter. This is actually quite similar to the second and third rhythmic reductions of Berlioz in his earlier *Symphonie Fantastique*. Thus, Saint-Saëns utilizes the fragment of the *Dies Irae* in a manner that radically varies from the original chant. In his reworking of the theme, he went even farther than either Liszt or Berlioz in divorcing the theme from its typical surroundings.

Saint-Saëns’ use of the *Dies Irae* is quite important as a later Romantic permutation of the medieval chant. However, it is really Saint-Saëns’ fusion of the medieval *Totentanz*, earlier Romantic influences, Goethe’s *Totentanz*, the concept of Death as fiddler, and the *Dies Irae* into a cohesive symphonic poem based on contemporary verses that offers such as fascinating example of the later Romantic’s lingering fascination with these ideas.

Just as in the medieval era, the revival of the *Dies Irae* and the *Totentanz* in music was allied with representations in the other arts. The macabre and frightening aspects of the medieval *Totentanz*, the *Dies Irae*, and the general idea of Judgment Day represented a captivating link with the past to Romantic artisans. None of these themes held the same power of admonition as during their original inception, but they were all a part of European heritage. Even as the Romantic artists manipulated and reshaped these ideas
for their own time, some eternal truths linked with these concepts remained. In particular, Goethe’s “Cathedral Scene from Faust and his Totentanz and Alfred Rethel’s “Scenes Auch ein Todtentanz aus dem Jahre 1848 and Death as a Cutthroat are significant examples of the lingering importance of these themes in Romantic art and literature.

Goethe’s realization of the Dies Irae occurs in the “Cathedral Scene” of Faust, where he interpolates segments of the text of the medieval poem. The “Evil Spirit” torments Gretchen in the cathedral as the choir sings the Dies Irae sequence. Yet another link between the “warning” of the Dies Irae is the fact that the scene directly following the “Cathedral Scene” is set during “Walpurgis Night,” a folk festival traditionally associated with witches’ orgies. So, here, as in Berlioz, the juxtaposition of the Dies Irae with a devilish rite or a witches’ dance is present. As earlier mentioned, Goethe’s twisting of the original Dance of Death theme into a midnight revel of skeletons was a purely Romantic ideal that inspired both Liszt and Saint-Saëns.

Final artistic exemplars of these topoi were the paintings and woodcuts of Alfred Rethel (1816-1859), a celebrated German artist. Specifically, an interesting permutation of the Totentanz is his woodcut series entitled Auch ein Todtentanz aus dem Jahre 1848, which encompassed six woodcuts meant to serve as a warning for the people revolting all over Europe. The entire series includes terrifying visions of Death as the propagator of revolution; he is really the only one who wins in a revolution, as he greedily consumes all of the classes. This series of woodcuts uses the medieval theme of the Totentanz and imbues it with fresh horror as a contemporary scene in the bloody revolts of the 19th century. Although many technological innovations improved the standard of living
during the Romantic era, the same technological breakthroughs led to increasingly ghastly casualties of European wars. Following the end of the Romantic age into the first World War, the immortal themes of the Dies Irae and the Totentanz were still a focus in European culture, but this time the art was driven by a true pessimism and despair in the ways of the world. Perhaps the Europeans of the 19\textsuperscript{th} and 20\textsuperscript{th} centuries had no better alternative to death and its ugliness than did their forefathers seven centuries before.

Epilogue

The two themes of the Dies Irae Sequence and the Totentanz have both enjoyed a long, intertwined history. Each evolved as different reactions to the universal human question of destination after death. However, the longevity and recycling of these mutual ideas throughout the history of art (of all kinds) is a testimony to their artistic worth. The Dies Irae and the Totentanz were not constantly quoted and used solely because of their subject matter; both are valid artistic triumphs of the human intellect. These two related ideas in turn generated more artwork as the centuries progressed. The subject of death as an eventual visitor to us all will probably forever maintain a central place in the iconography of art. Yet the concentrated use of both these medieval themes in 19\textsuperscript{th} century orchestral music, art, and literature is one of the great hallmarks of the century. Each reflects the fascinating social mores and ideals of the complex culture of 19\textsuperscript{th} century Europe. The use of the Dies Irae and the Totentanz in non-liturgical music mutually mirrored the European attitude, but also undoubtedly augmented these 19\textsuperscript{th} century masterpieces. The two medieval themes offered the works of Berlioz, Liszt, and Saint-Saëns both grounding in a rich artistic tradition as well as an aid to “Romanticize”
the eternal themes of death and resurrection. Perhaps it is through this Romantic
“mirroring” of these themes that we in the 21st century are still able to understand the
meaning of the Dies Irae and the Totentanz, albeit in surroundings drastically divorced
from their original milieu.
Bibliography


Delacroix, Eugene. “Lithographs of Faust.”

*Europäische Totentanz-Vereinigung.* http://www.totentanz.online.de/totentanz.htm


Goethe, Johann Wolfgang von. “Der Todtentanz.”


*Gospel of Matthew. The New American Bible.* Wichita, KS: Catholic Bible Publishers,


Abstract:

Traditional Holocaust studies have largely overlooked women’s unique voices, instead treating the eloquent and moving narratives of such renowned authors as Elie Wiesel and Tadeusz Borowski as definitive sources on “the” Holocaust experience. Recently, scholars have addressed the absence of women’s voices in Holocaust studies, arguing that women’s experiences, and their reactions to those experiences, were in fact very different from those of men. This topic is a controversial one, and some scholars argue that women’s suffering should not be focused upon in the context of an event that sentenced all Jews to death.

With such controversy surrounding this issue, the thesis of this paper is that works of imaginative literature and film offer a way to test whether women’s experiences should truly be held as distinct from those of men and, if so, what these differences were and whether they caused a profoundly different effect on women survivors. In short, were women “double victims” because of their gender?

Following the premise that women authors and directors might prove to be more likely to portray women’s unique experiences, this paper compares works of Holocaust literature and film by female authors and directors with a like number of distinctly male voices. This study pays particular attention to portrayals of what could be termed as women’s “double victimization,” such as the separation between mother and child, the mother’s frequent inability to save her child, and sexual humiliation and rape. Because of the sensitive nature of the types of victimization many women endured, this study determines whether each author or director has portrayed women’s double victimization sensitively, or whether it seems that women victims have been exploited for prurient interest.

While it seems that women authors and directors might have proven to be more perceptive of women’s double victimization, this paper reveals that some male authors and directors have proven remarkably adept at depicting women’s experiences effectively, yet sensitively. However, previously overlooked female authors like Charlotte Delbo and Cynthia Ozick can contribute greatly to a better understanding of women’s double victimization, often revealing new insight into Holocaust experiences that have been so widely documented by men.

This paper’s conclusion supports the arguments of scholars who claim that women’s unique experiences during the Holocaust are deserving of more study, while proving that the traditional canon of Holocaust literature and film cannot provide a complete understanding of the complex phenomena of victimization that occurred during the Holocaust. This study will become increasingly important as the literature and film of the Holocaust move further into the domain of popular culture, challenging audiences and artists alike to develop an understanding of and sensitivity to the double victimization of women.

Chapter One: Introduction: Double Victimization

Although Anne Frank, clearly the best-known author of the Holocaust, was a young woman, the fact is that most of those who have influenced our perception of the Holocaust have been men. Despite the many women who have contributed to the growing body of fiction, memoirs, poems, plays, and films about this period, the moving and eloquent testimonies of such men as Elie Wiesel and Tadeusz Borowski have come to be regarded as encompassing what all victims suffered during the Holocaust, and most studies have treated their narratives as definitive sources on “the” Holocaust experience. This very influence, however, has tended to mute the less strident voices of women authors such as Charlotte Delbo and Nelly Sachs.

Recently scholars have begun to address the absence of women’s voices in Holocaust studies, arguing that women’s experiences, and their reactions to those experiences, were in fact very different from those of men. Joy Miller is one scholar who believes that overlooking such distinctly feminine issues is to negate these women’s unique experiences. She writes that, “The thoughts, feelings, and perspectives of women in Auschwitz reveal distinctions unique only to females” (Miller 185). Myrna Goldenberg concurs with this statement, explaining that we must
closely examine the memoirs of women as well as men in order to represent the Holocaust more fully (327).

The subject of gendered differences in the Holocaust is a controversial one. Lawrence Langer is one Holocaust scholar who disputes the importance of gender during the Holocaust: "It seems to me that nothing could be crueler or more callous than the attempt to dredge up from this landscape of universal destruction a mythology of comparative endurance that awards favor to one group of individuals over another" (Preempting 58). The aim of scholars such as Miller, however, is not to award favor to women at the expense of men, but rather to include the often-overlooked experiences of women in Holocaust studies that have overwhelmingly focused on men.

There is no doubt that every victim of the Holocaust—man, woman, Jew, or gentile—was subjected to dehumanizing physical and mental torture. Women survivors, however, have depicted very different experiences from men in their autobiographical testimonies. In the case of Jewish victims, Joy Miller attributes these differences to the fact that women faced a "double jeopardy" of being not only Jewish, but Jewish women, the child bearers who alone had the ability to carry on the Jewish "race." Daniel Patterson is another scholar who holds this opinion; he writes that one unique aspect of the Holocaust was the "muder not only of human beings but of the very origin of human life and of human sanctity"—the murder of the Jewish mother and child (7-8). Indeed, it has been argued that the Nazis' "final solution" was one of the first such events in history that "did not treat the female population primarily as spoils of war but instead explicitly sentenced women and children to death" (Ringelheim, "The Split" 344).

To accomplish this end, women with children were generally selected for death upon arrival in concentration camps, while others were separated from their children or forced to make what Miller calls "choiceless choices" by selecting one child over another in an effort to save at least one from impending death (xxi). It was made a capital offense for women to become pregnant or bear children in the camps; women were often forced to kill their own infants, as well as other women's infants, in order to survive (Karay 298, Goldenberg 329).

The violation of the sanctity of the relationship between mother and child was not the only defilement of women during this extraordinary period. Some scholars cite experiences of women that are unique due to their biological sex as reasons to study the importance of gender in the Holocaust. Myrna Goldenberg highlights the vulnerability of young women as objects of potential sexual assault and humiliation (335). Indeed, many women were forced to prostitute themselves, faced sterilization and mandated abortions, and were subjected to degrading acts of sexual perversion. Miller writes that, although both men and women faced sexual and physical abuse, "women were subjected to atrocities that men rarely experienced or reported" (185).

While the sexual exploitation of women and the problems mothers faced concerning their children and child-bearing seem to be important reasons to acknowledge gendered differences during the Holocaust, some scholars argue that it is also important not to marginalize women further by focusing strictly on their uniquely feminine experiences. As Sara Horowitz writes, "Limiting our discussion in this way would—ironically—serve to reinscribe male experience as normative for the development of a master narrative, and would relegate women to the category of the mother, or the victim of sexual abuse" ("Women" 375). Indeed, many proponents of gendered studies of the Holocaust emphasize that gender is only one component of survivors' experiences, and that studying differences between the genders should provide better understanding of the victimization that occurred during the Holocaust (Ringelheim, "The Split" 349-350; Horowitz, "Women" 366-371). Horowitz continues this reasoning for a gendered approach to Holocaust studies by writing, "Their [women's] writing expands our cognitive and psychological understanding of the Holocaust, using narratives of victimization and survival to mediate on the problematics of memory, testimony, and trauma" ("Women" 374).

Joan Ringelheim explores these "problematics" by relating the story of "Pauline," who had been sexually molested by male relatives of the people who were hiding her during the Holocaust. Pauline was unsure whether her molestation was important within the context of the Holocaust. Ringelheim explains Pauline's insecurity about the importance of her own victimization by stating, "Her memory was split between traditional versions of Holocaust history and her own experience. [ . . . ] A line divides what is considered peculiar or specific to women from what has been designated as the proper collective memory of, or narrative about, the Holocaust" ("The Split" 344). Therefore, women victims like Pauline have been unsure whether their own testimonies of suffering belong with other "normative" experiences recounted from the Holocaust. With its absence, the missing voice of these women has greatly shaped our perspective of the Holocaust. For, as Ringelheim argues, though Pauline's story might not be typical, "If Anne Frank's diary remains the paradigm of hiding, we will never know, because it will be assumed that danger lurked only when Germans located those in hiding" (345).

The search for a better understanding of the complex phenomenon of victimization during the Holocaust is a common theme in arguments advocating the importance of studying women's unique experiences during that time. Scholars such as Lawrence Langer, however, are concerned that there is a danger in "overstating the importance of a biologically unique experience," arguing that the "ultimate sense of loss unites former victims in a violated world beyond gender" (Preempting 56-57). Langer is not the only dissenting voice from the host of scholars advocating gender studies of the Holocaust. His opinion is echoed by one Holocaust survivor, Ruth Bondy, who expresses...
her own misgivings by writing, “Cyklon B did not differentiate between men and women; the same death swept them all away. [ . . . ] Why should I focus on women? Any division of the Holocaust and its sufferers according to gender seem(s) offensive to me” (310). However, Bondy continues by stating that she did not want the story of the women of Theresienstadt, where she herself was imprisoned, to be “left out” of our knowledge about the Holocaust. Bondy tells the story of mothers of young children who had the opportunity either to present themselves for selection to work, or to go directly to the gas with their children: “Only two of about six hundred mothers of young children appeared for selection; all the others decided to stay with their children to the end” (324). Although the wish to remember the selfless actions of these women is reminiscent of arguments validating gendered studies of the Holocaust, Bondy reemphasizes her point of contention by writing that “Most of the prisoners[ . . . ], both men and women, tried to stay humane to the end, united as human beings” (325).

Joan Ringelheim notes that some could argue that rape, abortion, sexual exploitation, and pregnancy are always a potential part of women’s lives, and that the ubiquitous nature of these experiences causes them to be irrelevant within the context of such a cataclysmic event (“The Split” 345). This view is implicitly upheld by Cynthia Ozick, the author of a short story titled “The Shawl” (to be discussed later in this paper), which seems to deal with the “gendered” issue of motherhood and the trauma of losing a child. Ozick believes that, by emphasizing the importance of gender in the Holocaust, we are attempting to identify the Holocaust as something that did not happen to “just Jews,” but to women: their being Jewish becomes a mere detail. She writes, “It is not a detail. It is everything, the whole story. [ . . . ] The Holocaust happened to victims who were not seen as men, women, or children, but as Jews” (qtd. in Ringelheim, “The Split” 348-349).

Lawrence Langer seems to concur with this statement; he believes that, as victims, the men, women, and children who were murdered or who survived the Holocaust were not agents of their own fate and were thus unable to fulfill their traditionally gendered roles. Langer argues that women’s efforts to create for themselves a gendered role during the Holocaust were futile and that these efforts were in fact “mocked” by events beyond any victim’s control. Because Langer believes that gendered roles were impossible to maintain during the Holocaust, he concludes that the issue of gender cannot be considered important in the aftermath of the Holocaust (Preempting 49).

Joan Ringelheim, however, argues that women were indeed able to fulfill their traditionally gendered roles within the camps. Evidence gathered from her oral interviews with survivors indicates that women combated the pain of starvation by sharing recipes, that they altered the rags they were given to wear into more adequate clothing, and that they turned mutual isolation into relationships or surrogate families. Whereas many of the men in the camps concentrated on their own individual survival, many women fought desperately to save other women prisoners with whom they had formed an emotional bond; as one survivor said, “Women’s friendship is different than men’s friendship you see ... we have these motherly instincts, friend instincts more ... But that’s what was holding the women together because everybody had to have someone to lean on, to depend on. The men, no ... the men didn’t do that” (qtd. in Ringelheim, “Women” 250-251). Ringelheim concludes that women transformed their habits of raising children or their experience of nurturing into the care of their new, camp families - they became mutually supportive of each other, helping them to survive the dehumanization and hopelessness of the camps (248).

At times, however, it is apparent that women were not able to fulfill their traditional role as mothers or nurturers during the Holocaust, such as when some women attempted to save the lives of mothers by killing their infants at the time of birth. While this fact might seem to uphold Langer’s argument that, ultimately, women were unable to be “mothers” in the camps, it does remain important to question whether we should not, therefore, recognize the horror and trauma created by the Nazis in stripping away this inherent maternal role. Gizella Perl, a woman who served as a doctor in Auschwitz, was forced to kill many newborn babies, who, if discovered, would have served as death sentences for their mothers. As a mother herself, she felt extreme anguish with the death of each infant and her own inability to perform her traditional role as a caregiver: “I loved those newborn babies not as a doctor but as a mother and it was again and again ni) own child whom I killed to save the life of a woman” (qtd. in Patterson 17).

Despite controversy, it seems that progress has been made in recognizing and studying women’s unique experiences during the Holocaust. Cynthia Crane is one scholar who has delved into this controversial area; her recent book, Divided Lives, demonstrates the previously overlooked suffering of those Jewish women who escaped internment in the camps by being married to, or the daughter of, an “Aryan.” These women, who were called Mischlinge by Hitler to denote them as “half-breeds” or “hybrids,” faced a duality of identity - that of Christian and Jew, German and Jew - though many were not practicing Jews, but had actually been baptized as Christians (Crane 24-26). These women suffered through the deaths of many of their Jewish family members at the hands of the Nazis, yet they themselves were spared. The emotional pain and guilt that these women suffered cannot be ignored, for as one survivor states: “We shared the fears of those who failed to survive persecution, but we also had to endure the shame of having fared better than our fathers, our relations, our friends. We did not emerge unscathed” (qtd. in Crane 33).

The arguments of scholars such as Joan Ringelheim, Joy Miller, and Cynthia Crane have sparked new studies that consider women’s unique issues during the Holocaust. S. Lillian Kremer
completed a recent study of women's diverse experiences during this time as represented in imaginative literature, focusing upon American writings by and about women and comparing their stories with eyewitness testimonies. In her study, Kremer concluded that there were, in fact, apparent gendered differences in the suffering and response of women. She did not argue that one gender had suffered more than the other, or that the suffering of one gender was more tragic, but that women had been vastly under-represented in studies of Holocaust literature.

It is not only in studies of Holocaust literature that women seem to have been continually under-represented. Joan Ringelheim draws attention to the fact that the United States Holocaust Memorial Museum's Permanent Exhibition contains no conceptualization of women during the Holocaust, although it does contain segments dedicated to the victimization of Roma and Sinti, homosexuals, Jehovah's Witnesses, and political prisoners. It is not that she believes a separate section dedicated solely to women is necessary, because "the lives and deaths of women are too integrated into the entire picture to segregate their experiences" (Ringelheim, "The Split" 347-348). She does believe, however, that within the Permanent Exhibition it should be recognized and indicated, where appropriate, that women were victimized in particular ways.

Because women's experiences have continually been marginalized, much of the public's understanding of the Holocaust could be distorted or incomplete. For instance, one of possibly the most influential genres of Holocaust narratives, film, often represents women victims of the Holocaust as undiminished icons of physical beauty.

With the hitherto modest consideration given to women's experiences during the Holocaust, as well as the large controversy surrounding the topic, the thesis of this study is that works of imaginative literature and cinema offer a way to test whether women's experiences should truly be held as distinct from those of men, and if so, what these differences were and whether they caused a profoundly different effect on women survivors. In short, were women "double victims" because of their gender?

In the chapters that follow, unique works of prose and poetry by women authors and survivors are compared with a like number of distinctly male voices. This study is then extended to a number of popular European and American films. While the focus in each case is on women as victims, the broader view inherent in this study reveals interesting and in some cases surprising twists on the position of Ringelheim and Miller. Without assuming to end all controversy, the conclusion argues the importance of exploring beyond the traditional canon of Holocaust literature to approach a more complete understanding of the complex phenomena of victimization that occurred during the Holocaust.

Editor's note:
The length of Ms. Copeland's thesis precludes publication in its entirety in this journal. We have chosen to publish Chapters One and Five. Chapter Two: Imaginative Literature Written by Male Authors, Chapter Three: Imaginative Literature Written by Female Authors, and Chapter Four: Imaginative Holocaust Films Directed by Men and Women can be found on the Inquiry website.

Chapter Five: Conclusion: New Perspectives on Double Victimization

Joan Ringelheim and Joy Miller are two proponents of gendered studies of the Holocaust who argue that women were subjected to different victimization than were men, thus necessitating an acknowledgement of and more studies about women's unique experiences. However, Holocaust scholar Lawrence Langer and author Cynthia Ozick both argue against focusing upon the suffering of women in the context of an event that sentenced all Jews to death, whether man, woman, or child. With such controversy surrounding this issue, the purpose of this paper has been to examine the body of imaginative literature and film rising from the Holocaust in an attempt to discover evidence supporting or contradicting the argument that women's experiences should be held as distinct from those of men.

Studying fictional representations of women in imaginative literature and cinema has proven to be an effective way to test whether recent arguments for gendered studies of the Holocaust are well founded. Given the array of works by male and female authors and directors reviewed here, it seems that both men and women address in some way the gender issues discussed by Ringelheim and Miller, often supporting the thesis that women were doubly victimized by not only being Jews (or, for the non-Jewish victims, simply prisoners), but women. The previous chapters having demonstrated that imaginative literature in fact reflects women's double victimization, this chapter will further investigate how male and female authors and directors have dealt with these unique experiences, with the ultimate goal of expanding upon the standard perspective of Holocaust experience offered by such renowned authors as Elie Wiesel, Tadeusz Borowski, and even Anne Frank. An important issue to be discussed in this chapter, then, will be how male and female authors and directors have treated women's experiences: whether women have been depicted in ways sympathetic to their plight, or whether their victimization is exploited for prurient interests or commercial exploitation. Because male authors have established the paradigm for Holocaust experience, this chapter will also consider whether or not there is a profound difference between the portrayals of double victimization given us by male authors and directors and those given by female authors and directors.

Of those studied, the male authors Borowski, Wallant, Styron, and Ka-Tzemik, the female authors Delbo, Sachs, and Ozick, and the film directors Spielberg, Pakula, Lumet, and Cavani all portray different degrees and perceptions of what could be termed women's double victimization. For the purposes
of this paper, the types of double victimization studied were the violation of the maternal bond between mother and child and the sexual exploitation of women through rape and sexual humiliation.

While Elie Wiesel largely ignores the issues surrounding women in the Holocaust, his novel Night can form a foundation of men’s experiences from which we can measure that of women. For instance, whereas Wiesel describes the competitive and often ruthless nature of the men’s camp, where even fathers could not always count on their sons, Charlotte Delbo’s play “Who Will Carry the Word?” demonstrates the solidarity and support of women in non-relative groups such as those discussed by Ringelheim. Indeed, while a more in-depth comparison between these two works is not possible here, further meaningful contrasts can be found, portraying very different reactions of men and women to such universal forces within the camps as starvation, abuse, and disease.

In contrast to Wiesel’s Night, Tadeusz Borowski’s This Way for the Gas, Ladies and Gentlemen forms a comparison between the men’s and women’s camps. His narrator, Tadek, bluntly describes the worse conditions of the women’s camp, dispassionately relating the necessity of women to use sex as a commodity in order to obtain essentials like food and clothing. Tadek also deals with issues concerning the women in the experimental block, pregnancy, childbirth, interrupted motherhood and “the Puff.” In the short story, “Auschwitz, Our Home (A Letter),” Tadek implores his own loved one interned in the women’s camp to “try to grasp the essence of this pattern of daily events, discarding your sense of horror and loathing and contempt, and find for it all a philosophic formula. For the gas chambers and the gold stolen from the victims, for the roll-call and for the Puff, for the frightened civilians and for the ‘old numbers’.”

Tadek’s instructions, however, to discard horror and loathing and contempt, belie Borowski’s true wish that he might be able to “give an account of the fraud and mockery to the living - to speak up for the dead” (116). Indeed, Borowski seems to be successful in relating that awful fraud and mockery of humanity that was the Holocaust, depicting not only his own experiences as a somewhat privileged political prisoner through the narration of Tadek, but also an extent of women’s double victimization. However, while Borowski is often effective in imparting the horror of women’s experiences, he does not explore the traumatic effects of their victimization, nor does his narrator offer a woman’s perspective, rather he demonstrates women’s victimization as merely a way of life in Auschwitz. In this way, Borowski does not supply his women victims with a voice; though he describes women’s experiences, he does not engender sympathy for their particularly unique plight.

American writers Edward Lewis Wallant and William Styron focus more upon the implications of double victimization than does Tadeusz Borowski, each approaching the Holocaust in a similar way with their respective novels, The Pawnbroker and Sophie’s Choice. Both novels reflect upon Holocaust experience years after the fact, portraying the lastingly traumatic effects of experiencing such horrors of the Holocaust as a parent’s impotence in the face of his or her children’s deaths. Wallant, however, further dramatizes the issues confronting his protagonist, Sol, by depicting Sol’s wife’s sexual humiliation, and later, death. While Ruth is depicted as a beautiful woman in Sol’s flashbacks to happier times, the brief portrayal of her victimization in the brothel effectively eschews eroticism, emphasizing instead her hopeless state.

The two authors diverge in their depictions of the survivors, however, for Sol is a nondescript, paunchy man, who wears old-fashioned glasses and seems older than his age. Sophie, on the other hand, returns to being the object of intense sexual desire that she had been prior to her Holocaust experience. She is an iconic, beautiful heroine, one who is still abused even in post-Holocaust Flatbush by her lover. By making Sophie so absolutely erotic in her post-Holocaust victimization, author William Styron detracts from the tale she recounts to Stingo, coloring her experience with Stingo’s overwhelming erotic fascination with her. This is unfortunate, for Sophie’s dilemma of having to choose to save one of her children’s lives is a profoundly moving example of Joy Miller’s argument for women’s double victimization.

The director Alan J. Pakula remains faithful to Styron’s present-day depiction of Sophie as an incredibly attractive woman in his (1982) film version of Sophie’s Choice, although he is careful to de-emphasize her sexuality during flashback scenes of the Holocaust. Likewise, in the (1965) film version of The Pawnbroker, director Sidney Lumet closely follows Wallant’s formula, which successfully demonstrates double victimization as well as the resulting traumatic effects upon both victim and survivor. These two films represent a departure from what seems to be a tendency of film directors to depict beautiful women as arguably erotic in their victimization. Interestingly, a woman directs the film most notable for this transgression: Liliana Cavani’s (1973) The Night Porter.

The Night Porter is effective in that it straightforwardly depicts women’s unique vulnerability within the camp to rape and sexual humiliation, as well as the long lasting effects of such victimization. However, this film can be criticized for engendering voyeuristic interest in Lucia, a victim who is portrayed as undoubtedly erotic and even complicit in her own victimization. Lucia enchants the audience with her beauty, and the perversely fascinating relationship between Lucia and Max ultimately undermines the power of her terrible victimization.

Steven Spielberg has been similarly criticized for a latent eroticism in his depiction of women’s victimization; however, his portrayal of their victimization seems less voyeuristic than emphatic, in the controversial shower scene as well as the
eroticly charged scene between Amon Goeth and his Jewish housemaid, Helen Hirsch. Indeed, Spielberg’s intent for the scene between Goeth and Helen seems to have stemmed from a desire to illustrate the irrationality of Goeth’s belief in the Jewish temptress, for it is obvious that Goeth’s advances are frightening and repulsive to Helen, whose life, of course, depends upon Goeth’s continued fascination with her. It is important to note, however, that Jewish men are depicted very differently from Jewish women in Schindler’s List and can neither be described as attractive nor sexual creatures; indeed, there has been no criticism as to the exploitation of men in Schindler’s List, although they too are shown unclothed in camp selection scenes. This fact lends an ambiguous note to Spielberg’s true intentions for depicting what could be thought of as a disproportionate number of attractive women in Schindler’s List, although the emotional impact and sympathy engendered for the victims during these most criticized scenes often seem to outweigh what has been called voyeuristic in this film.

To return to literature, Holocaust survivor Ka-Tzetnik's House of Dolls also occupies an ambiguous position between portraying women’s experiences perceptively and forthrightly, and portraying their experiences voyeuristically. While this novel effectively imparts the horror of forced prostitution in a camp brothel, it risks attracting prurient interest for what is possibly the inherently voyeuristic nature of such a direct narrative about rape, mutilation, and sexual humiliation. Nevertheless, Daniella’s wish for her experiences to be remembered forms a litany throughout the novel; though the nature of her experiences might attract voyeuristic interest, the author’s motive for relating Daniella’s experiences seems to stem from a desire to inform others about such experiences that are today, and were then, incomprehensible.

Although Cynthia Ozick is not herself a Holocaust survivor, her two short stories, collectively titled The Shawl, are perceptive and sensitive in depicting women’s double victimization, intuitively realizing the disintegration of a mother’s sanity along with the death of her child. With The Shawl, Ozick deftly addresses women’s unique issues of violated motherhood and rape within the camp, illustrating the dilemma of a mother who has gone to great lengths to keep her child alive, only to become unable to help her at the most critical moment, in which the powers of the camp force this mother to repress the maternal urge to save her child from death. Seeing her screaming daughter being carried across the camp to be thrown against the electric wire, Rosa knows that any effort she might make to save her child would be futile and could only result in her own death, as well. In witnessing Magda’s murder, however, Rosa is forever changed. She is rendered unable to live with the reality of her daughter’s death, so she must fabricate her own reality in order to reconcile the fact that she lived, while her daughter died.

While Ozick is powerful in her illumination of women’s experiences, Rosa’s victimization through rape is merely implied; Magda’s true parentage is revealed to the reader only through innuendo and supposition. Rosa struggles, at times unsuccessfully, to suppress the horror of her memories of this victimization, although she finally admits, “I was forced by a German, it’s true, and more than once” (43). Ozick was, however, the only female author in this study to address the issue of sexual exploitation - Nelly Sachs’ two poems studied here focus entirely upon the destruction of the maternal bond, while Charlotte Delbo’s play focuses more upon the supportive relationships between women and the degrading effects of de-feminization in the camp. In fact, Liliana Cavani, director of The Night Porter, was the only woman studied who straightforwardly confronted this important factor of double victimization, although her film seems to emphasize the voyeuristic interest that can occur when dealing with this particular type of victimization.

Conversely, several of the male authors and directors, such as Ka-Tzetnik, Wallant, Spielberg, and, to a lesser extent, Borowski, straightforwardly explore the issues of prostitution, rape, and sexual humiliation, often very effectively demonstrating the horror of this victimization upon women. While one can only speculate as to the reason why the male authors in this study seem to be more willing to delve into this type of victimization, the inherent shame and degradation of such a traumatic event could affect the ability of women to expose it so openly. However, one can question whether it is indeed necessary to portray this victimization so explicitly, for works like Ozick’s The Shawl and Lumet’s film The Pawnbroker, both of which merely imply women’s sexual victimization, seem to be just as effective in imparting the very real trauma of such events as rape and sexual humiliation. Ka-Tzetnik and Cavani, however, choose to portray women’s victimization more graphically, raising the issue of whether their intent is to increase our understanding of the complex victimization that occurred during the Holocaust, or whether there is an underlying, prurient interest driving these depictions of attractive female victims in such an arguably exploitative way. Perhaps the extremely sensitive nature of this victimization automatically renders any attempt at realistic portrayal voyeuristic and even, at times, pornographic - an issue that is not easily solved. Despite this dilemma, the fictional representation of the Holocaust in poetry, drama, fiction and film continues with no apparent slackening of either audience interest or the readiness of artists to tackle this extraordinarily complex and freighted subject.

The extent to which male authors and directors dominated this literature is giving way slowly to a richer mix of men’s and women’s voices, especially in the memoir. Whether this new balance will produce works more sensitive to the double victimization of women is as yet unclear. The evidence of this study is that, while some male authors and directors have proven remarkably sensitive in this regard, previously overlooked female authors can contribute greatly to a better understanding of this victimization, revealing new insight and interesting comparisons.
to Holocaust experiences that have been so widely documented by men. Indeed, it is clearly evident that Elie Wiesel’s Night and even Anne Frank’s Diary can no longer remain the major narratives influencing our understanding of the Holocaust. While both of these works form important perspectives, they are only two pieces of the puzzle that makes up this incomprehensible period in history. However, as the literature of the Holocaust moves increasingly into the domain of popular culture, the challenge for audiences and artists alike will be to develop sensitivity to the double victimization of women without slipping into a kind of pornographic exploitation of female vulnerability. When that happens, the obscenity that was the Holocaust is compounded.

End notes:

‘See page 112 for reference. The “old numbers” were those camp inmates who had been imprisoned the longest, which often gained them seniority in the prisoners’ camp hierarchy. The numbers tattooed on their arms were much lower than newer arrivals’ numbers, and Borowski describes these older numbers as a source of pride for inmates because they had survived for so long within the camp.

Bibliography:


Sachs, Nelly. “0 the Night of the Weeping Children!” Schiiff, H. 69.


Filmography:


Faculty comment:

Ms. Copeland’s mentor, Professor Mark Cory, believes that her work is truly extraordinary. In his letter of nomination he wrote:

One of the deep satisfactions for me in teaching comes from the occasional crystallizing moment when a gifted but rudderless student discovers a way to focus a cluster of interests into a coherent pattern, which then becomes richly suggestive in terms of research and career. Almost two years ago, after working at a superb level in my course on literary reflections of the Holocaust and having visited the US Holocaust Memorial Museum in DC, Shauna found such a pattern. The thesis she defended at the close of the Fall semester represents the culmination of two years of continued reading, a successful SILO grant proposal, and presentations at two conferences. Shauna will be recognized this spring as an honors graduate summa cum laude for the totality of her work; this last piece

http://scholarworks.uark.edu/inquiry/vol4/iss1/1
is one of the most exciting cooperative projects I have participated in at this institution.

As an English major, Shauna has acquired excellent analytical and writing skills, which she combines with a more general interest in gender theory and its application to popular culture. Her dilemma coming into her final year was that she became rather more intrigued about the problems associated with the literature of the Holocaust than about the subjects typically chosen by English majors for their honors theses. Knowing then that we would be bringing three noted feminist scholars of the Holocaust to campus during the fall semester of 2001, I suggested she read the recent books by Cynthia Crane and Joy Miller to see whether an application of their provocative thesis to fictional women in Holocaust literature might be of interest. This was the genesis for what seems to me to be an absolutely cutting-edge topic. The controversy stirred by Crane and Miller, respectively, is whether women victims of the Holocaust suffered in a qualitatively different way than men. Noted scholars such as Lawrence Langer have been loud in their dismissal of this idea, almost strident. The most famous spokesperson for Jewish victims in this country, Elie Wiesel, himself a prolific author of Holocaust literature, is less dismissive, but has stopped short of subscribing to the thesis. Not surprisingly, Wiesel's fiction depicts the plight of male protagonists; women, to the extent they figure at all, serve merely to acknowledge and call attention to the male protagonists as objects of feminine desire. What had yet to be done, and what Shauna undertakes with her thesis, is to test the Crane and Miller arguments against some of the best-known female characters in the canon of Holocaust literature to see whether in fact authors have recognized and incorporated a qualitatively different kind of suffering in their depiction of women victims, and if so, whether there is yet additional difference in the way male and female authors articulate this different suffering.

In that fall I was able to introduce Shauna to Cynthia Crane, Joy Miller and Elaine Martin (Univ. of Alabama), and to invite her to the 2001 conference and workshop for educators held annually in Northwest Arkansas. The topic of that conference was perfect: “Women in the Holocaust.” By April of 2003, Shauna shared preliminary results of her study at the undergraduate research conference in Arkadelphia. She then won a Sturgis study abroad grant to travel to Europe and visit sites related to the Holocaust in Vienna, Berlin and Prague in order to gauge the extent to which women's voices are being featured in exhibits and memorials, if at all. This past fall she was a featured presenter at the 2002 conference and workshop for educators, reading a paper extracted from the thesis.

I see Shauna's work as marvelous preparation for graduate study, with the potential for exciting teaching and scholarship in her future. At the annual meeting of the German Studies Association in Washington, D.C., last fall, papers related to this controversy attracted more attention than any other. Her honors thesis committee found her work mature enough for publication, a judgment already anticipated by her two conference presentations.
GERARD DAVID’S NATIVITY TRIPTYCH: LANDSCAPE AS A GENRE AND A TOOL FOR SPIRITUAL PILGRIMAGE*

By Laurel Eddleman
Department of Art

Faculty Mentors: Dr. Lynn Jacobs
Department of Art

Abstract:

This work by Gerard David, painted around 1510 to 1515, represents a style built predominantly upon the foundation of his predecessors, while hinting in certain ways to the tastes and styles of the near future. Specific details of style show influence of artists before David such as Jan Van Eyck, Rogier Van der Weyden, Geertgen tot sin Jans, Hugo Van der Goes, and Hans Memlinc. His innovation is exemplified on the exterior panels, which introduce landscape as a genre and incorporate the spiritual theme of pilgrimage between the exterior and interior panels.

Stylistically, the Nativity shows a look toward the future and upcoming contemporary styles by exhibiting a hint of Italian Renaissance influence in the monumentality of the figures. The holy figures depicted seem to fill their space with a more structural quality than those of previous Northern paintings, which tend to lay figures on a surface with a more doll-like quality.

The unique and most revolutionary part of David’s Nativity occurs in the style and especially the iconography of the exterior panels. The exterior side panels, which are now separated from the central panel, show a full landscape completely void of figures, making this a first in the history of Netherlandish painting. Artists such as Geertgen tot sin Jans had begun to develop landscapes further by using isolated trees of relevant species to the scene, but most landscapes served only the purpose of a backdrop. Now David has indulged in the richness of full, green foliage, creating an intimate forest space. A small donkey and ox can be seen in the forest scene, making the only real connection to the interior where the ox and donkey are worshipping Christ with the others.

*Editor’s note: This paper received the University’s award for outstanding undergraduate research in the arts and humanities in 2002. The time necessary to secure permission to publish the paper’s illustrations made the paper’s publication in Volume 3 impossible, however.

This work by Gerard David (see figures 1, 2), painted around 1510 to 1515, represents a style built predominantly upon the foundation of his predecessors, while hinting in certain ways to the tastes and styles of the near future. The wings of the work are separated from the central panel (Metropolitan Museum of Art, New York) and held by the Mauritshuis, The Hague on long-term loan from the Rijksmuseum, Amsterdam. One who is familiar with early Netherlandish painting would never mistake this triptych for a category other than early Netherlandish. It contains all the elements of the Netherlandish style that had been in the process of creation for the century prior to the production of this work. However, David now at the end of a long line of masters and a master of the tradition himself, was in a position to appeal to growing tastes outside the established, setting up a unique dialogue between tradition and innovation. This element of innovation is most evident in the exterior panels of his triptych, which introduce landscape as a genre while integrating the spiritual value of pilgrimage with the other panels. Pilgrimage was an important and popular form of worship during this time, in the physical sense as well as a more individual, mental form of worship. David has uniquely utilized and capitalized on pilgrimage using a pure landscape to represent this notion of spiritual travel.

Stylistically, this painting fits into Gerard David’s later period (post-1511), when he had begun to gain freedom from extremely static compositions. In 1515, while still living in Bruges, David had become a member of the Guild of St. Luke in Antwerp, which had by then become an even more active and dynamic city than Bruges. The market here may have provided him with the opportunity to try to break new ground in style and message of paintings. Nevertheless, the

Nativity was done at the beginning of this period and is still characterized by relatively stiff forms and a duller palette than his very latest work.
David is undoubtedly indebted to Jan van Eyck (Van Schoute, 546), as are most painters after him, for his treatment of space and light, and his acute attention to detail, which became a trademark of early Netherlandish painting. In this

Nativity Triptych specifically, the brocades and pattern of fabrics and the detailed background viewed through an open window are reminiscent of Van Eyck. Several other works by David include clear references to Van Eyck as well. In his Cervara Altarpiece, Musee du Louvre, Paris, he includes over a scene of Christ’s crucifixion, a depiction of God the Father very much like the one in the Ghent Altarpiece. In looking at the details of the Nativity Triptych, I think there is slight evidence of Rogierian influence as well, as seen in the fluttering drapery of the angels’ garments. They float weightlessly like those found in many of Rogier van der Weyden’s Crucifixions (figure 3) and other works.

Other obvious influences were those from his native city Haarlem, including Dirk Bouts for his structured and well-defined spaces and Geertgen tot Sint Jans for his work in developing landscapes, (of which David became a great innovator). Specifically, Geertgen was probably key in influencing David’s hand at landscape, since he commonly went beyond the normal notion of landscape as consisting of standard shapes and textures in the distance making up the backdrop of a scene, and included landscapes with the motif of the isolated tree (figure 4). Some of David’s early faces also resemble those of Geertgen.

Hugo van der Goes can also take some credit for several of David’s compositional arrangements. An Adoration of the Magi done by David echoes clearly a composition of the same subject by Hugo, with its sharply receding wall and similar clumsy shepherds. In the Nativity, the very naturalistically rendered straw and basket in the foreground resemble the objects composed in Hugo’s Portinari Altarpiece, and especially of Hugo’s own Nativity in Berlin. David’s iconic representation of holy figures, very still and expressionless for the most part, may be due in part to his direct predecessor in Bruges, Hans Memlin. Memlin is known for his soft, quiet representations that one might say give his work a certain “Hallmark card” quality by modern standards.

In addition to these qualities of rich tradition in David’s Nativity, there are hints of innovation in the interior panels and a rather revolutionary exhibition of landscape on the exterior wings. On the interior, and especially the central panel, the figures are rendered in a monumental style, occupying real space instead of appearing flat on a plane. The faces and bodies are more modeled, as if they might be able to be viewed in the round, and are not quite so crisply handled, giving the scene a softer, more atmospheric quality. David has also created a real space for the birth of Christ with the crumbling, but still sturdy and well-defined structure in which the birth has taken place. These characteristics point to an attempt to assimilate some Italian values into the work, since the Italians at this time were studying anatomy and proportion to achieve a more real depiction of space and the human figure.

While the interior factors are subtle evidence of David’s transition away from the traditional, the most obvious and revolutionary element of innovation is found on the exterior side panels (figure 2). These panels of the triptych display a rich landscape void of any figures, making this a first in Netherlandish painting. David creates a deep, intimate, and quiet forest scene by covering most of the surface with naturalistic foliage, and leading the eye of the viewer between the trunks on the forest floor. Stylistically, we see a break from the traditional Eyckian execution. Instead of painstakingly executing the details of each leaf, he uses a slightly more painterly hand and creates the overall effect of full green foliage. Infrared reflectography and X-radiographs show underdrawings that laid out the basic composition of these panels on the ground layer with a light brush sketch. The trunks, foliage, and sky space were blocked out - apparently somewhat differently than the finished painting exhibits (Ainsworth, 243). Somewhere in mid-painting, David eliminated some trunks and closed in more foliage over the sky, creating the intimate and inviting forest space that now exists. David demonstrates a great knowledge of color in creating the naturalism he has in his forest. He uses a dark green for the background foliage, a middle mixed green for the closer branches, and a sunny yellow for the closest branches in full light (Ainsworth, 244), creating the realistic and inviting piece of forest.

Though the exciting change of style exhibited on the exterior of David’s Nativity was a step away from tradition, the real revolution occurred within the subject matter itself. This innovative approach to landscape – one that was more than just a backdrop – became a precursor to a genre that would become increasingly popular in later sixteenth and seventeenth century Flemish art. Up to this point however, landscape had been limited to playing second fiddle. It served the purpose of creating a location for a figure or set of figures, but was never appreciated for its own beauty. Before David, as mentioned before, Geertgen tot Sint Jans specialized minimally on the landscape by including individual trees instead of distant forests or mountains. It appears that he sometimes represented localized, recognizable species of trees relevant to the subject of the painting, indicating that he probably worked directly from nature at times. (For example, his Saint John the Baptist in the Wilderness (figure 4) uses isolated trees of varieties thought to be common in a Middle Eastern climate.) (Ainsworth, 217) David was surely influenced by this technique; however, with this Nativity he became the first to paint a landscape without any trace of a human figure, making it a significant turning point in Flemish art.

Not only is this landscape a statement in itself, but it also seems to have another meaning as it is combined with the rest of the painting. At this time the issue of travel and religious pilgrimage had become a popular theme in religion as a result of
the writings of a Carthusian monk of the twelfth century, Ludolph of Saxony, who wrote

_La Vita Christi_ (Life of Christ). This and many other popular religious writings described the life of Christ as a series of events - joys and sorrows, and encouraged the reader to participate in these events of the life of Christ personally. Many people were reading these types of writings and even taking pilgrimages themselves. However, most commonly, people used these books and accompanying images of pilgrimage to perform a spiritual pilgrimage in their own mind. They were encouraged to picture mentally Mary holding the baby Christ, crying in her arms on the road to Egypt, or to see the drops of sweat and blood that dripped from Christ's brow as he carried his cross, and to walk this path with him.

It seems that David has utilized his revolutionary landscape to serve as a conduit for a mental and spiritual pilgrimage. Within the forest scene, he has created a path that winds intimately between the tree trunks (Ainsworth/Christiansen, 290). As a viewer, the naturalistic landscape draws you in to a space where you can actually exist, and feel the serenity of the meditative mood created. As you wander through and around the trunks of the tall trees, you come to a house or inn, just as Mary and Joseph did as they sought a suitable place for the birth of the Son of God. Perhaps this is where they were refused a room. Barely recognizable, an ox and donkey lie and graze peacefully, while silently alluding to the stable scene inside. The forest scene is serene and quiet. It asks for thought and invites meditation. So they - the Holy Family, and you - the viewer, continue on until you reach the interior, where the manger is found, according to prophecy, and the joyous birth happens.

The triptych format was extremely popular (and practical) at this time because it allowed for several different scenes to be presented - using both interior and exterior space. In David's _Nativity_ Triptych, the format lends itself effectively to the idea of pilgrimage because it allows you to "travel" logically from exterior to interior on a consistent path. The exterior here plays a traditional role in the dialogue of the triptych as a whole (Ainsworth/Christiansen, 279) - it introduces the scene inside. Instead of the usual Annunciation scene to introduce the Nativity however, the landscape is allowed to lead us there. Although it is disputable, most recognize that the existence of the ox and donkey in the forest scene on the exterior panels are used as a connecting device to the interior panels. Maryan Ainsworth claims that contemporary viewers would have easily recognized the ox and donkey as a reference to the Nativity scene. (Ainsworth, 211)

In light of this emphasis on the journey of life (Christ's), there are several references to travel on the interior panels as well. Joseph is unusually depicted as a young man in traveling clothes, which includes a short robe, soft shoes, a cloak, and a walking stick. The basket in the foreground may be a traveling basket since it is depicted in many paintings on the subject of the Flight to Egypt. David himself did at least two versions of the _Rest on the Flight to Egypt_, which was another theme elaborated upon in the _Vita Christi_ with an emphasis on landscape and travel. These common symbols of travel would have also have been easily recognized by the viewers of the sixteenth century. (Ainsworth, 211) Other symbols in the central panel emphasize the importance of Christ's purpose and allude to his ultimate crucifixion, another step in the journey of Christ's life, and part of the pilgrimage of the faithful. The dilapidated building is a common symbol of the life and passion of Christ, symbolizing the passing of the Old Law into the New Covenant. A dandelion grows out of a crack in the crumbling building. Since the dandelion blooms at Easter time, it is a symbol of grief in reference to Christ's sacrifice. (Ainsworth/Christiansen, 283) The sheaf of wheat placed prominently in the foreground of the painting, and the basket of clothes next to it both point to the sacrifice of Christ's crucifixion by referring to the wheat as "the body of Christ" in relation to the Eucharist, and the clothes as burial wraps. All of these symbols would have been used to focus the mind of the viewer on the image of Christ.

The donors of the _Nativity_ would have been familiar with the traditional symbols for Christ and may have been particularly interested in the ideas of pilgrimage and travel, since those symbols are less common. They may have expressed a desire for David to convey these interests accurately, since he is known for his ability to understand and cater to the requests of his clients, even outside his key subjects and styles (Metropolitan Museum of Art). The donors are probably meant to serve as examples of the faithful to those who would view the triptych. They are presented here in the garb of saints Catherine and Anthony (characterized by their wheel and pig, respectively), although their exact identity is yet to be determined. Saints Jerome and Leonard, who may have been personal patron saints or those of a church or chapel of the donors, accompany the two donors to the holy scene. The implication is that they have completed the pilgrimage and are now partaking in the joy of Christ's birth, hopefully encouraging others to do the same. In addition to a didactic role, the portrayed donors have a rather self-serving purpose: to establish not only their own piety, but also their status and wealth. They are slightly larger than the Holy figures of the central panel, and are depicted in a space closer to the viewer. Many times the donor, not the artist, would determine many of the specifics of a commissioned painting, which offers rationale for their prominence in many Northern Renaissance paintings.

Despite the personal and social ambitions evident on the interior wings, this triptych is ideal for displaying the message of the call to follow the journey of Christ. The relationship of the iconography of the exterior and the interior makes David's _Nativity_ most intriguing - the exterior is used to present a mood and introduce a story, and the interior is the satisfaction and joy of coming to the destination of a journey. David has invited the
viewer to join the pilgrimage in a whole new way. His delicate, but satisfying landscape calls to enchant the viewer’s mind, intending to lead the way to spiritual meditation.

While some have called Gerard David a ‘tardy phenomenon 4’ for his continuation of tradition in a changing artistic world (as well as economic, social, and religious), he is to be commended for his innovative approaches to incorporating his style with contemporary iconography and novel subject matter into effective and appealing paintings. He has received the honor of painting the first pure landscape, which obviously became a noteworthy claim soon after his career when landscape became a popular genre. But perhaps most interesting is the way he has used this innovation as a medium for taking the viewer on a spiritual pilgrimage to the birth of Christ. Whether the viewer was looking on the painting as a fresh creation or is seeing it now as a piece of history, this triptych is remarkable in its attempt to lead the viewer on the path of worship using the beauty of fresh landscape and the inspiration of traditional religious imagery.

End notes:
5 Grove Art. [www.groveart.com/data/articles/art Gerard David, Life and Work (Bridgeman Art Library source)]
7 James Hall, Dictionary of Subjects & Symbols in Art (Boulder, Colorado, 1979)

Bibliography:
Grove Art. [www.groveart.com/data/articles/art Gerard David, Life and Work (from Bridgeman Art Library source)]

Figure 1: Gerard David, "The Nativity with Donors and Saints Jerome and Leonard." The Metropolitan Museum of Art, The Jules Bache Collection, 1949. (49.7.20 a-c)
Figure 2: Gerard David. "Two Forest Scenes." Royal Cabinet of Paintings. Mauritshuis, the Hague.
Figure 3: Roger van der Weyden, "Crucifixion." Vienna, Kunsthistorisches Museum
Figure 4: Geertgen, "John the Baptist in the Wilderness." Staatliche Museen zu Berlin. Preussischer Kulturbesitz Gemaldegalerie.
Abstract:

Waterfronts have played an important role in the settlement of North America. Not only did waterfronts provide protection and security to newly forming settlements, but these waterways also provided the opportunity for trade while acting as a major hub of activity and social interaction. Ironically, the waterfront that was the genesis of the city was often neglected after the introduction of the steamboat and the railroad as well as industrial advancements. Polluted and neglected, riverfronts of the cities were in disrepair. Not until the past few decades have larger cities revisited their waterfronts and realized the amenity that lies at the edge.

Van Buren, Arkansas, though not a big city, struggles with the same abandonment of its industrial foundation. The Arkansas River with its origin in Leadville, Colorado, has created a riverfront for the city of Van Buren as well as a historic Main Street. As in many other places, the city has turned its back on the potential of Van Buren’s riverfront. The industrial facets of a poultry plant, railroad lines, and a ten-foot high levee wall segregate the city from the riverfront.

The metamorphosis of the Van Buren Riverfront involves creating a mixed-use riverfront development that not only provides opportunities for growth but also serves as an icon for the city. By acknowledging the existing site conditions and respecting the industrial character therein, Van Buren can reclaim the riverfront area as a vibrant edge.

Concept:

The concept that drove the design of the riverfront was derived from the existing architectural character of the downtown historic Main Street. Victorian architecture begins the descent down Main Street, followed by an area of governmental buildings, which include a juvenile detention center and the county courthouse. Transitioning from a governmental significance to an industrial nature, the street is bisected by two lines of railroad tracks that act as a switching yard. Proceeding over the tracks and under the trusses of the Simmons Poultry Plant, Main Street is terminated by an ignored riverfront park. Conceptually, the Van Buren riverfront can be seen as reminiscent of a butterfly. A butterfly, often ignored in its earliest stages, develops into a beautiful attraction of uncompromising attention. The Victorian architecture of Main Street is the pupa of development from which the larva or governmental transition takes place. Wrapped in the chrysalis is the mystery of what the industrial area could become. Through metamorphosis of this plan, the resulting beauty of the riverfront has been given wings.

Just as in a butterfly’s metamorphosis, Van Buren has gone through changes in appearance, character, and function—changes that cannot necessarily be seen taking place; but, like a butterfly forming within a chrysalis, the change is happening.

Goal:

Like other larger cities that have revitalized riverfronts, Van Buren is perfectly poised to reap benefits of reconnecting with its riverfront. A riverfront master plan began with the realization— that downtown Van Buren needs to acknowledge the amenity at its edge. The master plan would include: a mixed-use riverfront development to provide opportunities for growth and development; a much-needed reconnection of the riverfront with downtown; and elements to serve as icons for the city and to promote its appreciation. Capitalizing on the riverfront’s location relative to downtown mixed-used development would allow for a connection necessary to sustain and enhance the downtown area while creating an urban green space. Establishing a connection with the riverfront to the surrounding downtown location would foster an appreciation for both the built and natural environment. Embracing the industrial aspects of the site and varying its architecture and function—would turn initial site constraints into opportunities for recognition of the riverfront environment.

Program Development:

To make the Van Buren riverfront separate and unique, the program elements within this design would not only include what other riverfronts already had but also build on what they did...
not have. In what ways could Van Buren’s riverfront development be given what other riverfronts had and more? How could people be able to interact safely with the water and experience its power and at the same time know its peace?

Precedents allow designers to see the successes and failures of things that have already been done. Although each city should be considered independently, based on its individual needs, precedents also give invaluable insight as to future needs for growing cities. Aker Brygge in Oslo, Norway, an international waterfront, was chosen as a precedent based on its sound mixed-use development. On a smaller scale, Central Riverfront Park in Cincinnati, Ohio, was the precedent selected based on its sense of place in association with the river. Louisville Riverfront Park in Louisville, Kentucky, on the other hand, was once an industrial site that has been revitalized and today has great visual appeal. These projects were chosen as precedents in an effort to synthesize the best possible design solution for the waterfront revitalization in the city of Van Buren.

Aker Brygge is an international waterfront that is worthy of exploration. Once a major industrial shipyard, the mixed-use redevelopment—now attracts an estimated six million people annually with its wide variety of cafes, retail shops, and office space. The development also houses two movie theatres, a kindergarten, and a decorated harbor-side entertainment site. The design is not only a combination of brilliant architecture, but also an intelligent planning of spaces that is sensitive to people while maintaining the primary focus on the waterfront. It maintains intimate spaces that relate well to the scale of the buildings. Aker Brygge, which is close to downtown and to the city’s historic City Hall, is designed to respect its connection to the older part of the city while introducing an exciting new architectural style.

Central Riverfront Park in Cincinnati, Ohio, was another case study that was investigated. Just as the Arkansas River helped to establish Van Buren, the Ohio River played a major role in the settlement of Cincinnati and its subsequent development. Central Riverfront Park’s proximity to the downtown Main Street and the seasonal flooding the park undergoes make this park comparable to the Van Buren riverfront. The park design strives to create a riverfront setting that acknowledges the river and the influence it exerts over the surrounding landscape.

Another precedent that was studied, Louisville Riverfront Park in Louisville, Kentucky, is a great example of the revitalization of an industrial site. Warehouses, barge facilities, and junkyards once lined the water’s edge. Today the city has taken back its waterfront. The riverfront park is 80 acres of environmentally sensitive parkland that has been creatively engineered to provide flood protection. The park has won popular acclaim because of its great lawn used for games and concerts, its festival plaza, its children’s play area, and its sculpted linear park.

The study of each of these waterfront areas resulted in individual insights as to possible program elements that could be potentially beneficial to the Van Buren riverfront. The Aker Brygge waterfront demonstrates the positive potential of a mixed-use development. To increase the number of people using the riverfront, there must be a variety of different uses—uses that are both active as well as passive—all the while maintaining the primary focus that attracted the users in the first place, the water. Study of Central Riverfront Park in Cincinnati, Ohio, contributed to the development of an icon for Van Buren. Central Riverfront Park paid homage to the Ohio River through interpretative and historical program elements. Van Buren needed to harness the power of the river in a similar way, a unique way. The industrial nature of Louisville Riverfront Park prior to its revitalization was comparable to the present image of Van Buren. Could a compromise be made between industry and nature, and could the industry be made to be beautiful?

Looking to each of the precedents for program development made obvious the need for a mixed-use development in a close but reverent relation to the river: a riverfront that would reveal not only the power of the water but the attraction at the water’s edge while maintaining architecture that was appropriate to both new and historic uses of the site.

Constraints into Opportunities:

As part of the design process, both opportunities and constraints must be considered. An initial site constraint is the poultry-product industry of Simmons Incorporated. The Simmons Incorporated plant spans across Main Street, blocking the view of the river and making entry to the existing riverfront park difficult. Since Simmons International comprises buildings that are of an industrial nature, there is an opportunity to build upon and enhance the existing architectural image. Changing the function, improving the appearance, and improving the visual character of the plant would increase the viability of the riverfront by attracting users.

Another site constraint is the two railroad tracks as well as the idle railroad lines that divide the site. Unfortunately, there have always been negative associations to the railroad with regard to safety. For example, a 1920’s film, “Teddy at the Throttle,” portrays a notorious image of railroads that often comes to mind. A woman has been tied to the tracks in the path of an approaching train and is waiting for a hero to rescue her. Society says we can interact with subways, buses, metro stations, etc.; but when it comes to trains, we perceive danger. The railroad tracks within this particular site are connected to a transition zone—a track-switching yard—so trains travel within the site only at a very slow pace. The riverfront development will create opportunities or people to see the function, build upon the appearance, and dissociate old ideas of negativity with maintaining a healthy respect for railroads.
The levee wall that also doubles as the historic wall mural poses a problem for the riverfront connection. While the intent is to retain the Arkansas River in the event of flooding, it also acts as a major divide between the city and the riverfront. The riverfront can be seen and accessed only by a fifteen-foot-wide entry through the wall. Sculpting the riverfront park with landforms that would act as a levee as well while functioning as a detention basin in times of flooding would allow removal of the levee wall.

**Vertical Ideas:**

Joseph Beuys speaks to all by saying, “Man is only truly alive when he realizes he is a creative, artistic being,” and that “even the act of peeling a potato can be considered a work-of-art if it is a conscious act” (Fineberg, 1940). The Van Buren site needed a defining feature—something that would tie the entire site together as well as capture the power of the river; consequently, the works of various artists were considered. Because the site already had such a strong industrial nature and the need for an icon was prevalent, industrial artists and more radical artists were the primary focus.

The works of Bernd and Hilla Becher, a couple whose black and white photography focuses on industrial and domestic structures such as gas tanks, coal silos, blast furnaces, grain elevators, preparation plants, oil refineries, framework houses, and water towers as their primary architectural subject matter, were utilized for inspiration. Christo and Jeanne-Claude, another collaborative couple, have done many projects that are inspirational as well as extraordinary. Christo and Jeanne-Claude enjoy using real objects such as oil barrels, bottles, or buildings. Stacking them, wrapping them, or packaging them, Christo and Jeanne-Claude take the ordinary and create something enticing, giving the objects new life.

Intriguing as well as inspirational, Bernd and Hilla Becher capture radical existing structures within the landscape that we have grown to accept aesthetically. On the other hand, Christo and Jeanne-Claude take radical approaches within the landscape to attract the viewer. In the work of Bernd and Hilla Becher, various shapes and sizes are the result of form following function; and, although the resulting function is not always portrayed in the most beautiful form, one can envision the radical Christo and Jeanne Claude taking these various forms and wrapping them in creativity.

A catalyst of ideas from these artists inspired the design process. One such idea was that of an industrial icon that would respond to the industrial character of Van Buren. Twenty solitary wind turbines, known as the Middelgrunden Wind Turbine Cooperative, the largest offshore wind farm, stand firm in the Oresund Channel between Denmark and Sweden and became the icon source. Not only are these industrial wind turbines beautiful, but they are also functional and their purpose beneficial. Although there would not be sufficient wind to power such a structure in Van Buren, other similar alternatives were considered.

**The Design:**

Conclusions were made after consideration of site constraints as well as evaluation of several design alternatives. In response to the existing grid and the topography of the site, the master plan is a continuation of that grid with strategically-placed nodes of activity that allow for both visual and physical connections to the riverfront. Designed to target future growth and expansion in all directions of the site, the master plan also utilizes the power of the Arkansas River and is sensitive to the floodplain within the riverfront park (Figure 1).

The initial site concerns in the design of the master plan were the existing buildings and the floodplain. The Simmons, Inc. buildings would be maintained, but their function would change to that of a mixed-use area, specifically aimed toward attracting people. The riverfront park, on the other hand, would be maintained as an open, urban green space capable of retaining water in times of flooding. Development on the river would be of minimal scale because of the necessary elevation changes that would have to be made. Elevating structures within the riverfront park might take away from the river and create a further separation from the natural environment.

The central node within the design, the marketplace plaza, has been developed in response to the major node on Main Street—the railroad depot/chamber of commerce (Figure 2). From the depot, an intriguing view to the marketplace plaza has been established by a semi-transparent tensile structure held in place with steel cables. The placement of this node allows traffic to penetrate all the way to the riverfront, benefiting existing businesses along the way.

Bordering the marketplace plaza to the southwest and bounded by the Arkansas River is the secondary node to the marketplace plaza—the riverfront amphitheatre (Figure 2). The organic shape of the amphitheatre with the moveable stage island was formulated so that boaters can actually come into the riverside area to experience staged events. The terracing of the amphitheatre allows people to interact with and also allows staged events to occur even during times of increased water elevation.

Another major node within the design is the Jefferson Street Bridge (Figure 3). The bridge offers excellent views of the downtown area as well as the happenings within the switching yard—views that cannot be appreciated from a vehicle. Because other streets already establish a way over the railroad tracks, it was decided that the Jefferson Street Bridge should become a pedestrian-only bridge. A pedestrian bridge would provide opportunity for mixed-use development on the bridge to attract people from the downtown area and gradually pull them to the river’s edge. The bridge, providing the best vantage point to view
the downtown area and the river icon, would become a major linkage between downtown and the riverfront park.

The end of the Jefferson Street Bridge literally supports the secondary node that consists of sculpted landforms that cradle a skate park (Figure 3). Wire cables span from the end of the Jefferson Street Bridge into the park to evoke the tension that has given shape to the landforms, all the while mimicking the movement of a caterpillar on a leaf. The skate park design is unique in that the landforms have been sculpted in such a way as to allow onlookers the safety and comfort of an elevated viewing platform. The landforms wrap around to the opposing side of the Jefferson Street Bridge and skate park where a bridge has been constructed to continue and complete the curve of the existing Jefferson Street Bridge. Cables from the continuation of the Jefferson Street Bridge anchor to the earthwork and mirror the same struggle of the opposite feature. This bridge is programmed to celebrate the Arkansas River with a barge ride that makes a circuit up and down the river, allowing passengers to experience the water-sculpture garden.

The water-sculpture garden consists of water mills, similar in concept to a windmill or a wind turbine. These two-bladed icons, staggered three deep, complete the visual axis of the (Figure 4). Powerful and vast, much like the existing industrial structures in Van Buren, these icons demand attention and attract people to the riverfront edge. The electricity generated by these icons, in combination with the current of the Arkansas River, powers the component of the water-sculpture garden that functions at night.

At night, the other element to the sculpture garden comes to life. The night component is a lighted sculpture shaped like the tail of a shooting star. Proper engineering keeps large, illuminated pieces of slag, a bi-product of glass, weighted just below the surface of the Arkansas River, allowing the riverfront to become a spectacle that can best be viewed when traveling across the bridges that connect the cities of Van Buren and Fort Smith. Contemporary, industrially-designed buoys connected by steel cables keep boaters at a safe distance from the water garden but do not create a visual barrier.

Complementing the water-sculpture garden are the curbside drainage feature and railroad plaza water features. Together, these three features act as a unifying element that ties the entire site together. The curbside drainage feature with removable grate collects storm water runoff, but also acts as a light feature at night. The same slag material used in the water-sculpture garden allows the drainage feature to glow at night, providing an icon of visual interest as well as a way-finder to the riverfront park. The railroad plaza comprises a promenade of cafes and shops running parallel to the riverfront. The railroad plaza provides an area of interest with a seat wall that acts as a misting-water feature reminiscent of the steam from old railroad engines. The plaza also contains a light feature at a safe distance from the railroad and allows visual interaction throughout the site (Figure 4).

The tertiary node slightly to the northwest of the marketplace plaza is intended to spur the opportunity of growth in that direction. This sculpted landform acts as another entry into the park allowing access for that side of downtown. Another half-bridge mimics the continuation of the Jefferson Street Bridge. This bridge as well is programmed to celebrate the Arkansas River and provide an overlook onto the children's playground that is cradled within the sculpted landforms.

Phasing:

In order to allow development to occur financially, phasing would be necessary. If phased properly, each phase of development would economically prepare for the next phase. Within each phase, the major nodes establish connections from downtown to the riverfront.

Phase one, the catalyst for the riverfront development, was established at the terminus of Main Street in response to the major node atop Main Street-the railroad depot and Chamber of Commerce. It is a mix of places of entertainment, with the marketplace plaza and amphitheatre nearby, as well as retail shops, restaurants including a cafe/coffee shop, a hotel, and loft apartments. The architecture is open and interactive within this phase, allowing views to the river as well as the train plaza. The sidewalks are busy with shoppers and outdoor eateries.

Phase two sets up further development of retail shops, restaurants, and loft apartments. Major development of the riverfront park, including the skate park and under-the-bridge parking, takes place. The levee wall is removed and replaced with sculpted landforms that act as a detention basin in times of flooding. A trail weaves throughout the riverfront park at an elevation that offers views onto all the happenings within the park and onto the open green spaces. The Jefferson Street pedestrian bridge, with under-the-bridge parking, plays a major role in this phase, providing the connection back to downtown and offering the best views of the site, including the Arkansas River. The bridge is programmed with a lookout tower as well as restaurants and some retail establishments.

Phase three completes the other half of the riverfront park but still allows for park expansion within this particular area. Development along the riverfront is much like that of phase two, but it also programs arts and cultural development as well as commercial opportunities. In addition to parallel parking throughout the site, another parking lot has been placed here to support growth within this area of the site.

Phase four makes the project complete, providing the icon for day and night view. The water mills and illuminated water-sculpture garden make this phase a stand-alone attraction. Because the icon will attract people, the two half-bridges have been
designed for views as well as for leisure programs — barge rides and fishing areas—that celebrate the river.

**Conclusion:**

The riverfront redevelopment master plan changes the relationship of the riverfront to the downtown. Not only does it reconnect the two and serve as a welcoming front door, but it also does so with the added benefit of increased potential for economic development and people presence. In doing so, it accomplishes the primary goal of bringing people back to the riverfront—not just for special events, but continuously through the new residential opportunities in the adaptive reuse of existing structures. Further, the riverfront redevelopment master plan creates opportunities for spaces and activities that will serve residents and visitors while creating an icon for the city. The metamorphosis of the riverfront has been given wings.

**Bibliography:**


Figure 2. Detail of Marketplace Plaza and Amphitheatre.
Faculty comment:

One of Ms. Gregory's faculty mentors, Mark Boyer, writes:

It is not uncommon for our students to choose senior projects that related to their hometowns—the territory is familiar and it gives students a chance to give back to their communities. It is, however, unusual to have a student take a hometown project as far as Juana did in hers.

Juana wanted her project to be the catalyst for enhancement of the Van Buren downtown. Juana used our standard design and planning processes to accomplish this project. After deciding that the type of project she wished to do was a riverfront redevelopment, Juana sought out precedents across the country and Europe to research the "state of the art" of practice in this arena. This search even led Juana to undertake a two-week intensive tour of riverfront redevelopments in Sweden, Denmark, and Norway to learn what approaches are being used for success.

Returning home, Juana set out to inventory and analyze the elements of Van Buren's downtown and surroundings. Through this work, Juana found both inspiration and opportunity to apply some of what she had discovered in her research about riverfront redevelopments. Her conceptual idea of metamorphosis was strong and derived from what she experienced in Van Buren. She was able to look at Van Buren in a new light and see elements that could be used as design inspiration and to create an identity for the city. She generated several design alternatives and, with the help of the faculty, evaluated those alternatives and chose one to develop further. Juana used this riverfront redevelopment to solve problems, create opportunities and build connections between the riverfront and downtown Van Buren.

While Juana utilized standard design and planning approaches to complete her project, she did not simply apply standard solutions. What makes her project unique is that she allowed the sense of place that already existed in Van Buren and the inspiration of her research to speak simultaneously to generate the solutions to problems and venues for under-utilized opportunities. Through Juana's design development, she was able to accomplish the necessary flood protection and create a vibrant and inviting connection between the new riverfront development and the existing downtown. Her work represents the highest level of critical and creative thinking—a level that we would like for all of our students to exhibit. It shows a meritorious ability to research and translate that research into appropriate application to the current design problem. This project exemplifies the capabilities of landscape architects and the role we can play in creating places that make a difference for people and communities.

Juana's ability to work from the scale of planning to site design to detail design also serves as a standard to which other students should strive. She was constantly coming to me with a new idea of how something she found could be incorporated into her development or how some detail could be a further explanation of the conceptual idea. Juana did not stop at designing the riverfront for Van Buren. She watercolored the final presentation drawings and made two models when only one was required. These two models were working models where she poured concrete and purchased slag to demonstrate the curb and gutter lighting system, and welded steel elements together to provide an image of the windmill icons. She incorporated music and video clips in her PowerPoint presentation to convey both what she saw in Van Buren and her vision for the riverfront redevelopment. She went on to apply her design concept to the design of her project report. She even went so far as to design special pages on which to print the documents and design and build special containers in which to house them. This level of excellence sets Juana's work on the pinnacle of student submissions.
Abstract/Problem Statement: Dar Islam Mosque

In historically Muslim countries, mosques take the prescribed form of a large hall with adjacent courtyard, minaret(s), and ornately decorated entrance portal. With Islam's spread to the United States, the mosque no longer takes this form, due to construction and technological conventions, as well as the diminished economic will of the religious community. The design for the Dar Islam Mosque in Albuquerque, New Mexico takes a position of difference in response to the current debate in the Muslim community over historicism versus contextualism of mosques, where the individual's response to the space is primary, with emphasis placed on disconnecting from the every day and forming a spiritual connection.

The mosque brings with it a set of rules, regarding the separation of the sexes, preparation for prayer, and the act of prayer, while the site brings another set of influences, including American culture, the desert climate, and the topographic nuances of the existing site. The challenge lies in mediating the needs of the mosque and the building's performance as shelter, tempering the separation of the sexes with the equality demanded by American culture, and giving to the mosque, which is rooted in certain ideas of faith and spirituality, a sense of the spiritual, where the typical form of a mosque is not mimicked or repeated, but where a visceral response or connection is made between the individual and the sacred.

The process of architectural design constitutes the research program, and the final design documents represent the discoveries of the following research process: (1) intense study and analysis of a cactus as an exemplary biological system; (2) generation of a wall system as interface between interior and exterior, using the cactus as precedent; (3) intense study of the Islamic faith and historic mosques; (4) field research; including examination and interpretation of the site and exploration of two existing mosques in northeast New Mexico; (5) investigation of site sequence and perception through a series of perspectival drawings; and (6) comprehensive design of building and site, including intense studies of enclosure, structure, heating, cooling, ventilation, lighting, and surface materials.

This comprehensive project, completed as a pinnacle experience in the design studio sequence, serves as a cohesive finale, integrating the principles presented in the architectural technology courses with the design process, while preparing the graduate for the multifaceted role of intern architect. More particularly, through challenging the use of imagery and symbolism in religious architecture, the research focuses on those things fundamental to the act of worship, where the individual's multisensory perception of the spaces is primary, with perceived visual, spatial, tactile, audible, and relational connections manipulated, challenging one's perception of reality - from tangible to impalpable - eliciting an association to things beyond the physical.

Mosque: a Phenomenological Investigation of Sacred Space

The luster and gleam of the stone, though itself apparently glowing only by the grace of the sun, yet first brings to light the light of the day, the breadth of the sky, the darkness of the night. The temple's firm towering makes visible the invisible space of the air. Heidegger

The design of a mosque for the mesa overlooking the city of Albuquerque, New Mexico, comes about as the penultimate project in the five-year design-studio sequence. The project seeks to integrate principles presented in technology, history, and theory courses with the design process, and the following paper analyzes the outcome of the project created with an understanding of Islamic culture and religion and an appreciation for the methods of phenomenology.

Sacred architecture has developed in unison with that of all other institutions; the role of the building has become one of utility and economy favored over human inhabitance. Contemporary architecture is a product of function and economics; buildings are constructed in an assembly-line fashion, becoming a creation of habit. "Habit dulls the mind so that a man builds with little more awareness of choice than does an animal that constructs instinctively," and the human roles in the environment,
human needs for dwelling, are ignored. Many paths have been and are being explored in reply to a need for modern architecture that responds to human needs, particularly the need for sacred spaces to worship, but I will focus on a phenomenological approach, in which an architecture attuned to the human experience is sought, in contrast with current solutions to the problem of habitation.

The architecture of the mosque is rooted in tradition. The components of a mosque spring from a prescribed set of forms with a prescribed set of functions. The mosque began as a simple place of worship, as seen in the prophet's house at Medina or in the Great Mosque of Isfahan, and evolved into institutions such as the Washington, D.C., Islamic Center. Nowadays, the size of the mosque denotes the number of faithful who come to worship, and the amount and quality of surface adornment attests to the wealth, power, and faithfulness of the patron or ruler. The simple act of worship has melded with political influence, social hierarchy, and control.

The Islamic faith forbids the use of icons; yet certain components of the mosque building and surface adornment become iconographic symbols when removed from the spatial patterns and cultural nuances of traditional or historic settings. For example, the intended function of a minaret is to elevate the muezzin to a great height so that his voice carries a considerable distance when calling the faithful to prayer; yet, with the advent of broadcasting systems, the minaret, which creates "identity" in an Islamic building through traditional association, has been diminished to a sign equivalent to the logo of a corporation. The minaret's intended function is again diminished when the mosque is taken out of an urban environment and placed within America's suburban sprawl and when there is no longer a captive audience of like-faith people but rather a complex mix of many faiths. As with the minaret, Arabic inscriptions in the manner of those adorning the Dome of the Rock lose their potency and meaning when they are used to adorn a mosque whose patrons do not speak Arabic. When these religious "icons" are applied in the diverse North American landscape, they are reduced to symbolic decoration, void of meaning and function but steeped in history.

The Islamic community is split between historians who believe that the traditional form of the mosque and traditional ornamentation must be used in order to prevent a departure from and possibly opposition to orthodox tradition, and those who believe the form a mosque takes should be determined by the community it serves. While my project reflects the latter belief structure, it must also confront a greater question: Can a building of completely modern design elicit responses equal to or greater than those of an ancient mosque? Several lines of inquiry arise from such a proposition: Is a phenomenological approach to mosque design an appropriate method for seeking alternative lines of inquiry into design and meaning in religious architecture? What does it mean to inhabit a sacred space, and how does a religious building help one to better understand his or her place within the religious collective and the world? How does the design for the Dar Islam Mosque address the question? Does the design for the Dar Islam Mosque fulfill both the pragmatic and spiritual needs of the faithful?

Phenomenology is a twentieth-century philosophical movement that studies the complex relationship between person and world, without recourse to theory, deduction, or assumptions from other disciplines. Pertinent overlaps exist in philosophy, psychology, and architecture that explore the very essence of what it is to be human and to exist in the world. These in turn form a platform through which the ramifications of being and dwelling in architecture and of architecture on being can be investigated. When Heidegger speaks about the Greek temple in "The Origin of the Work of Art," he emphasizes the power of architecture to make the world visible. Architecture has the power to disclose things, to make them appear as what they are. He says, "The temple in its standing there, first gives to things their look and to men their outlook on themselves." This is the point of inquiry, where traditional methods of building and inhabitation are questioned. With a new sense of piece and space-making and awareness of the power of the building to form deeper visceral connections with the inhabitant, the building attains an elevated level of presence. The person may not know why he or she is drawn to that place or why that place evokes the thoughts and feeling it does, but the building has an evocative presence that establishes itself in memory as an image, a feeling, or a relationship.

To design a building within a phenomenological framework means that the architect must think about every surface, every plane, every opening and about how every part of the building enhances one's awareness of his or her being, so that one is more conscious of the surrounding and how he or she relates to it. Nevertheless, even when designing without phenomenology in mind, one can, as Heidegger does with the Greek temple, analyze the project from a phenomenological stance, looking for those instances that connect, that form a dialogue with the inhabitant, heightening his or her awareness of being. Analysis of the formal and experiential aspects of the Dar Islam Mosque reveals the influence of this way of knowing.

As one approaches the mosque, the visitor crosses a threshold formed by a grove of desert apricots. This threshold is the embodiment of difference; it delineates the boundary between everyday routine and sacred reflection, while forming a connection between the tree as a creation of God and man as a similar creation. The skin of the building forms additional but different connections. It captures the presence of the sun; the building reflects the light, and the shadow reacts to the light. The skin is also an indicator of both weather and time. The surface is Corten metal sheet, which, when exposed to the elements, weathers, forming orange rust deposits on its surface. The skin of the building also guides water over its surface, channeling it along a circuitous path, holding the water longer in some places,
and allowing the water to be slowed and directed to specific points. In turn, the water leaves on the skin a remnant of its presence in the form of rust (except where the skin folds inward, protected by the overhanging metal above). The water that rushes over the building’s surface is collected in the courtyard and is represented in the ablation chambers, where the water washes over the body, leaving its mark on the individual and suggesting to the psyche its cleansing and life-giving qualities.

The carving away of the earth and the placing of the building within it rather than on it evoke a sense of its being of the earth yet other than the earth. The earthen wall, juxtaposed against and facing the man-made wall, frames the earth, the sky, and, in the distance, a volcano now dormant but once a raging fire within the earth which forced up the mesa on which one now stands. Basic elements—earth, air, fire, and water—are brought together in this place, interacting and enhancing one’s self-awareness. A specialized space of transition, anticipation, and preparation is carved into the earthen wall, adjacent to the prayer hall, where one sits in a slick, clean, dark room, lit from above by a thin shaft of light, which cleanses the mind while it cleanses the skin as a part of the ritual of prayer. The thin strip of light that washes the concrete wall dances across the water as it flows over the skin and falls to the floor. The light makes the darkness appear, evoking a somber mood, as it emphasizes the cleansing, life-giving water. The hard surfaces of the chamber increase the resonance of the water as it drips or flows, focusing the attention away from competing sounds of vehicles or conversation, clearing the mind, and preparing one both mentally and symbolically to enter the prayer hall.

Inside the prayer hall, the temperature is much warmer. Both light and materials influence the temperature. The warmth of light, reflecting off the warm, rusted Corten, imbues the space with a transcendental feeling, much like the light in a church or cathedral. The light is not static, but the openness of the wall changes from floor to ceiling. The wall contains no apertures at the floor, and the number of openings increases as one’s eye moves up the wall to the roof. This draws the eye upward away from the immediate, increasing one’s awareness, while shifting the focus from the individual to the spiritual.

The design for the mosque in Albuquerque explores alternate methods for design. Typically, the plan generates the design process; here, it was considered last. The use of vignettes took precedence and drove the process. The feeling of the place or space was recorded through perspective drawings and sketches before the place was realized in traditional forms of architectural expression—i.e., plan, section, and elevation. This design process allowed for innovation both in a way of working and in a finished design. The existing site, the ritual, the procession, and the feeling of the spaces created became primary over the programming of space. This design does not try to emulate the past and is not dictated by the lack of economic will or technological and construction conventions seen in many religious structures being built today. Moreover, the mosque seeks to form visceral connections, not through icons or signifiers, but through consideration, manipulation, and amplification of human consciousness and self-awareness.

End notes:
3 Yi-Fu Tuan, Space, and Place, (Minneapolis: University of Minnesota Press, 1977) p.
4 Salim A. Elwazani, "Sacral Qualities of Form in Mosque Architecture: Transformation of the Arts at the Qur’an into the Arts of the Mosque" on IslamOnline.net
5 Qur’anic passage, 5.92: “O Believers, wine and arrow shuffling, idols and divining arrows are an abomination, some of Satan’s work, so avoid it; so haply you will prosper.”
7 Akel Ismail Kahera, Deconstructing the American Mosque: Space, Gender, and Aesthetics, (Austin, TX: University of Texas Press. 2002) p. 65.
11 When referring to temperature, I am using the fourth definition given by Merriam Webster: relative state of emotional warmth.
Faculty Comment:

Ms. Harlan's faculty mentor, Professor Marlon Blackwell, was lavish in his praise of her project. He said:

Melissa's proposition for the vertical surface is an intelligent tectonic system that accepts anomalies and eccentricities as it synthesizes structure, environmental systems, and lighting, and yet, simultaneously has the capacity for allowing exquisite spatial qualities to emerge in the project at a variety of scales - the scale of the site, the scale of the mosque and the scale of the individual. It is a project deeply engaged with its time and place; richly articulated in sequence, material, and joinery, she has devised a well-crafted and philosophically coherent structure.
Figure 8. Prayer hall interior

Figure 9. Exterior wall section

Figure 10. Axonometric

ARCHITECTURE: Melissa Harlan. Dar Islam Mosque
SECTION II: SOCIAL SCIENCES AND BUSINESS

ECONOMICS, FINANCE, PSYCHOLOGY, AND HISTORY
A FREE TRADE AGREEMENT OF THE AMERICAS:
A CASE STUDY OF BRAZIL

By Maria Eliana Cadario
Department of Economics
Faculty Mentor: Dr. Tracy Murray
Phillips Petroleum Distinguished Professor of Economics

Abstract:
This paper will examine the likely impacts of the proposed Free Trade Agreement of the Americas (FTAA) that was initiated by President Bush in 1994 and is anticipated to include 34 countries in the Western Hemisphere excluding Cuba and to come into force during 2005.

The paper will focus on the likely trade effects for Brazil, the largest potential member of the FTAA. In the first part, I will review the current trade relations between countries in the Western Hemisphere and the US, including the various bilateral and multilateral agreements such as NAFTA, MERCOSUR, CBI, The Andean Trade Preference Act, etc. Next, the paper will present the special complexities introduced by rules of origin that are inherent in any free trade area.

The methodology used to estimate the trade impacts is presented in section 3. Briefly, the elimination of US tariffs on imports from Brazil will stimulate US imports to the benefit of US consumers and at the expense of US producers and imports from other countries. Standard comparative static analysis will be used for the base estimates. These estimates will be qualified for special situations, namely the US quotas on sugar imports, the extremely high US tariffs on orange juice imports, and the forthcoming change in the world trading environment for textiles and apparel.

The results are that the FTAA will provide significant benefits for Brazil, and by implication, for the other Latin American countries that will also benefit from the FTAA.

1. Introduction

This paper will examine the likely impacts of the proposed Free Trade Agreement of the Americas (FTAA) that was initiated by former President Bush in 1994 and is anticipated to include 34 countries in the Western Hemisphere excluding Cuba and to come into force during 2005. The paper will focus on the likely trade effects for Brazil, the largest potential member of the FTAA. In the first part, I will explain what the FTAA is and I will describe the trading arrangements under this agreement. In the second part, I will review the current trade relations between countries in the Western Hemisphere and the US, including the various bilateral and multilateral agreements such as NAFTA, MERCOSUR, CBI, The Andean Trade Preference Act, etc. Next, the paper will present the special complexities introduced by rules of origin that are inherent in any free trade area.

The methodology used to estimate the trade impacts is presented in section 3. Briefly, the elimination of US tariffs on imports from Brazil will stimulate US imports to the benefit of US consumers and at the expense of US producers and imports from other countries. Standard comparative static analysis will be used for the base estimates. There will be calculations of trade creation and trade diversion for the major dutiable products that are currently traded between the US and Brazil. These estimates will be qualified for special situations, namely the US quotas on sugar imports, the extremely high US tariffs on orange juice imports, and the forthcoming change in the world trading environment for textiles and apparel.

The expected results will be that the FTAA will provide significant benefits for Brazil, and by implication, for the other Latin American countries that will also benefit from the FTAA.

1.1 What is the proposed FTAA?

The Free Trade Agreement of the Americas (FTAA) is a proposed agreement among the economies of 34 countries in the Western Hemisphere, from Canada to Chile. This agreement excludes Cuba, The FTAA seeks to eliminate barriers and to increase trade and investment flows. This agreement would eliminate tariffs between FTAA countries within 10 years. It would also eliminate regulatory barriers that restrict trade. The Free Trade Agreement of the Americas is not a new idea that has emerged from recent negotiations. The dream of a unified American Continent reaching from the Article Circle to Tierra de Fuego inspired statesmen and thinkers of both North and South America decades ago. -J/

During 1960s, the Latin American countries began regional integration as a means of accelerating their development. Latin
American countries signed multilateral agreements creating some of the trade pacts that bind together the North with the South. Some of these early agreements are: The Central American Common Market (1960), Latin American Free Trade Area (1960) and Andean Pact (1969). After these trade agreements were created, the intraregional trade within Latin America increased, but the economic and political crises that erupted in the 1970's brought regional integration to a standstill.

The decade of the 1980's was not easy for hemispheric relations. Latin American countries struggled to cope with the worst economic crises since the Great Depression and the United States was dealing with the civil wars in Central America. These years were later designated as the "lost decade" for South and Central America because of the lack of growth due to the negative effects of the debt crisis. Nascent democracies in Argentina and Brazil were hit by hyperinflation. Civil strife burdened societies throughout Central and South America.

During the past decade, Latin American countries have improved considerably. These countries have implemented development strategies that combined macroeconomic stabilization policies with increased trade and regulatory reform. Privatization programs have reinvented important sectors, especially transport and utilities. New capital, technology and management skills have been introduced. Another factor that has helped these countries has been the provision of substantial funds to both pay debt and increase infrastructure investments. The objective behind this strategy was to encourage investment from both domestic and foreign sources to stimulate economic growth. As a result, international competition for investment funds has pressured governments in Latin America to accelerate their reforms. Faster growth in these economies created new opportunities for trade and investment. Domestic economic reforms continue to propel the integration process in the Western Hemisphere.

At the same time, regional integration arrangements have evolved and economic and political ties have solidified between countries in South America, Central American and Caribbean communities. These regions have established custom unions among the partner countries and they are also expanding their trade ties with countries in other regions in the Hemisphere. The agreements that have been developed range from simple tariff reduction pacts to comprehensive free trade agreements and custom unions. For example, MERCOSUR, the Southern Cone Common Market (Argentina, Brazil, Paraguay and Uruguay) is strengthening its custom union and has entered into free trade negotiations with Chile, Bolivia and other countries in the Andean Community. In addition, Mexico and Canada have concluded free trade pacts with Chile and Costa Rica; Mexico has agreements with other Central American neighbors, Bolivia, Colombia and Venezuela. Recently (2002), Chile has signed a free trade agreement with the United States. These countries must now adapt quickly to changing conditions in the world markets in order to gain market share and investment funds. Countries cannot use protection barriers to safeguard their industries from foreign competition. Countries must use their regional pacts to reinforce domestic reforms and to prepare themselves to compete effectively against foreign firms at home and abroad.

After acknowledging these pros from economies integration, the Free Trade Agreement of the Americas was launched by many representatives of the hemispheric countries. This FTAA platform comes from the initiative of the North American Free Trade Agreement (NAFTA) and the Enterprise for the Americas Initiative (EAI). NAFTA was the first reciprocal free trade agreement to link a developing country as an equal partner with developed countries. The EAI was the first initiative to link trade, investment and debt issues in a coordinated approach to economic development in Latin America. The creation of the EAI came from two concerns that the United States had about the Latin American countries. First, Latin American countries needed new inflows of foreign capital because their debt problem was not going to be solved without prolonged economic stagnation. Second, the economic and political reforms in this region would not be possible without accelerated growth. The United States concluded 14 agreements on trade and investment with hemispheric countries in 1990-1991. These agreements were the foundation for later negotiations of more comprehensive pacts such as NAFTA. NAFTA was born from the recognition that closer ties with neighbor countries can create great opportunities. NAFTA held the argument that it was going to benefit not only economic growth but also some political issues such as promoting democracy in Mexico and contributing to a long-term solution to immigration problems.

In December of 1994, the United States offered to host the Summit of the Americas in Miami. The Presidents of the Americas met to set a new path in the relations among Western Hemispheric countries. Although the Summit addressed several issues, trade and integration were the centerpiece and the Free Trade Agreement of the Americas was proposed. The Miami Summit initiated a detailed work program to prepare for the negotiations of the FTAA. The Declaration of Principles was: "Although faced with differing development challenges, the Americas are united in pursuing prosperity through open markets, hemispheric integration and sustainable development." The 34 presidents committed their governments to begin the construction of the Free Trade Area of the Americas. The representatives set 2005 as the deadline for the conclusion of negotiations. In March 1998, the FTAA negotiations were finalized in the San José Declaration of trade minister. The following month, the summit leaders reconvened in Santiago, Chile and officially launched the hemispheric trade negotiations. The Santiago Summit set the agenda for the FTAA. There were twelve negotiating groups at first. They were created to address market access and a consultative group on small economies was developed to ensure that the
concerns of the majority of the FTAA countries were reflected in the work of each group. Responsibility for the organization of the FTAA process was given to the Minister of Trade (the United States) while the Vice Ministers of Trade were accountable for managing the preparatory process. This process consisted initially of meetings of trade ministers called Trade Ministerials (the first one in Denver 1995 and the most recent in Quito, Ecuador November 2002). The objectives of these meetings were to define the issues to be addressed in the negotiations and recommend to the presidents when the official negotiations of the FTAA might begin. The negotiations began in Miami in 1998. The talks were set up so that the big and small, the rich and the poor countries would share the responsibility to carry on the negotiations. The United States and Brazil were assigned the responsibility to co-chair the final stage of the negotiations from November 2002 to the end of the talks.

The Declaration of San Jose drafted the objectives and terms under which the FTAA negotiations need to be conducted. These objectives are:

- To promote prosperity through increased integration and free trade
- Establish a Free Trade Area that eliminates barriers to trade gradually, in areas of goods, services and investments no later than 2005
- Market openness Incentive integration of the smaller economies into the FTAA
- Make trade liberalization and environmental policies mutually supportive
- Secure protection of worker rights

There are also certain principles that guide the negotiations among partner countries in this agreement:

- Decisions are made by consensus, so each country has veto power
- Transparency
- Consistency with rules and regulations of the World Trade Organization
- Commitment to improve on WTO rules and disciplines
- Single understanding with simultaneous negotiations in all areas ("nothing is agreed until all is agreed")
- Coexistence of the FTAA with bilateral and sub regional agreements
- Countries negotiate and take responsibility for actions individually or as members of sub regional groups
- Special attention to smaller economies and difference in levels of development
- Rights and obligations shared by all members
- Countries need to make sure that the national laws conform to FTAA obligations

The main purpose of the FTAA is to promote growth and prosperity of the member countries by eliminating barriers to trade and investment. It is clear that the FTAA will not exist as a final agreement until each issue has been negotiated with the approval of the 34 countries. The FTAA negotiations cover three main areas: market access reforms, including liberalization of trade barriers and removal of discrimination against foreign suppliers in the application of domestic law; rules covering trade and investment in goods and services sectors; and trade facilitation measures.

The twelve original working groups of the preparatory stage became nine Negotiating Groups, a Consultative Group and two special committees for the FTAA negotiations. Each has a Chair and Vice Chair. These positions rotate to obtain geographic balance. The nine groups are divided in areas of main interest:

1. Market access issues for goods
2. Agriculture
3. Services
4. Intellectual property rights
5. Subsidies, antidumping and countervailing duties
6. Government procurement
7. Investment
8. Competition policy
9. Dispute settlement

The other advisory committees were created to deal with (a) problems that arise from the participation of small economies, (b) inputs from representatives of civil society and (c) Internet use and difficulties presented in electronic commerce in the hemisphere. There is also the Administrative Secretariat that supports the negotiations. It provides logistic and administrative services; translating and interpreting services and managing the official documents. The Inter-American Development Bank (IDB), Organization of American States (OAS) and the United Nations Economic Commission for Latin America and the Caribbean (ECLAC) compose the Tripartite Committee. This committee provides technical support and its participation during the preparatory stages has been crucial. As mentioned above the nine negotiating groups are the main keys to achieve effective negotiation in this agreement.

Editor's Note: Ms. Cadario's paper includes, in sections 1.2 through 1.14, descriptions of the roles various commit-
tees, groups, and authorities play in negotiating and implementing free-trade agreements. Space precludes publication of these sections here. The complete paper is available on the Inquiry website.

2. Current bilateral and multilateral Agreements in the Western Hemisphere

It is important to emphasize that the FTAA will not be a substitute for the current regional arrangements in the hemisphere. Those pacts will coexist and complement the hemispheric agreement. The readiness of Latin America and the Caribbean countries depends on the reforms that those regional agreements undertake. It is necessary to have success in those "small-scale" trade pacts to ensure that the FTAA will be successful too. Regional trading rules may be changed to comply with hemispheric-wide standards but in other cases when regional trade rules go beyond those in the FTAA, the regional rules would not change.

2.1 North American Free Trade Agreement (NAFTA)

The members of this agreement are Canada, Mexico and the United States. It is a free trade area of 387 million people. The total intragroup trade is $437.8 billion. US imports from Canada and Mexico are $230.2 billion. Exports to the US equal 83% of total Canadian and Mexican exports. NAFTA came into effect in January 1994. This agreement was created to promote free trade in goods and services and increase investment. After the US-Canada agreement, NAFTA has tried to improve government procurement, intellectual property and investor rights and it has created more stringent rules of origin. NAFTA also eliminates non-tariff barriers such as import licenses and, guarantees fair and open competition. Unlike MERCOSUR or the ANDEAN Pact, NAFTA does not have a common external tariff to nonmembers. NAFTA has improved the trade relations among the member countries. The main trading partner of the United States is Canada. Mexico is also getting benefits out of this deal, by exporting more goods into the US at cheaper prices than before. According to the Council of the Americas, the trade of the US with Mexico and Canada accounts for one-third of all US merchandise trade and it exceeds the trade that the US has with Europe and Japan. As a result of NAFTA, US accounts for 76% of Canadian imports and 74% of Mexican imports. Mexican products have also entered the US market strongly and these effects contribute to the focus of the US to expand trade with Latin America.

2.2 Caribbean Community and Common Market (CARICOM)

The members of this custom union are Antigua & Barbuda, Bahamas, Barbados, Belize, Dominica, Guyana, Jamaica, Montserrat, St. Kitts & Nevis, St. Lucia, St. Vincent and the Grenadines, Suriname and Trinidad and Tobago. The representative market is of 6 million people. This agreement came into effect in August 1, 1973. The objective of this agreement was the eventual integration of its members and economies, and the creation of a common market. CARICOM has never been effectively completed. A fully implemented common market would significantly enhance the market potential of these countries. For CARICOM, a hemispheric agreement can provide stronger trade relations and investment links with North and South America to further expand their economies. Even tough CARICOM countries do not account for a large portion of the industrialized countries trade; these small countries could improve its trade volume and expand economically. One of their problems is that members have similar export products and similar economies (bananas, hotel occupancy, tourism, fishery, etc).

2.3 Caribbean Basin Initiative (CBI)

The members of this agreement are Antigua and Barbuda, Aruba, Bahamas, Belize, British Virgin Islands, Costa Rica, Dominica, Dominican Republic, El Salvador, Grenada, Guatemala, Guyana, Haiti, Honduras, Jamaica, Montserrat, Netherlands Antilles, Nicaragua, Panama, St. Kitts and Nevis, St. Lucia, St. Vincent and the Grenadines, Trinidad and Tobago, and the United States. This is a preferential trade arrangement. CBI imports from the US is $14.2 billion 29% of total imports; US imports from CBI countries is $14.7 billion. Exports to the US equal 36% of total CBI country exports. The objective of this agreement is to revitalize the economies of the Caribbean countries trough tariff preferences. CBI was established in 198 US eliminated duties on al products except textile and apparel products, canned tuna, footwear, certain leather good and certain watches and watch parts.

2.4 US- Andean Trade Preference Act (ATPA)

The members of this agreement are Bolivia, Colombia, Ecuador, Peru and the United States. The ATPA has a common external tariff. There is a representative market of 103 million people. It was implemented in May 1988. The objective of this agreement is to establish a free trade area with a common external tariff and eventually become a full common market. Another objective is to expand economic alternatives for Andean countries and to combat drug production and trafficking. The United States is the main export partner for the Andean Pact countries. This has become a fast growing market for US exports too. The Andean Pact has a common external tariff (CET) ranging from 5, 10, 15 and 20 percent. The Andean group has intended negotiations with MERCOSUR to create a free trade area that joins together the two trading blocs. It would take some more time to agree to a free trade area among these countries since a study done by the Inter-American Institute for Cooperation on Agriculture found that the agricultural conditions among countries of ATRA and MERCOSUR are largely different.

2.5 Southern Cone Common Market- Mercado Comun del Sur (MERCOSUR) or Mercado Comun do Sul
(MERCOSUL)

This is probably the most important agreement that involves countries in Latin America. The members are Argentina, Brazil, Paraguay and Uruguay. Bolivia and Chile are associate partners. It is a custom union with a representative market of 200 million people. It came into effect in January 1995. The objective of MERCOSUR is to establish a common market in the south that is expected to be fully implemented at an unspecified date. MERCOSUR has a combined GDP of $1 trillion (approximately two thirds of the GDP of South America). MERCOSUR imports from the US $ 18.8 billion, 22% of total imports, US imports from MERCOSUR is $ 11.4 billion. Exports to the US equal 15% of total MERCOSUR exports. All tariffs will be eliminated and MERCOSUR members will fully integrate with a common external tariff schedule. The maximum tariff rate will be 20%, with an average tariff rate at 10%. Individual national rates apply to imports of capital and high-technology goods until next decade when capital goods receive a CET of 14%. In 2006, a CET of 16% will be applied to informatics and telecommunications. Chile signed a free trade agreement with MERCOSUR in October 1996 and Bolivia did the same in March 1997.

The trade relationships with the US started in early 1990’s with talks about trade and investment framework agreements. MERCOSUR started better than expected because of the economic reforms undertaken by the two biggest players: Argentina and Brazil. In the 1990’s, both countries ended their hyperinflation. Programs of privatization also fueled an increase in foreign investment in these two countries. However, the recent economic crisis in Argentina has put MERCOSUR in a standstill situation. In order to sustain economic growth, the countries of MERCOSUR will have to integrate their markets even more by improving transportation links, speed up custom processes and sustain macroeconomic stability. With the potential for growth and economic benefit, MERCOSUR represents an important trading bloc for the FTAA. Brazil’s trade with the US is primarily in manufactures while other countries seek reform of agricultural trade barriers and subsidies. According to Brazil’s ambassador to the United States, in the year 2000, Brazil had a weak performance in exports to the United States because of discriminatory treatment due to NAFTA, competition from the rest of the world, and US import restrictions. For Brazil, the FTAA represents good opportunities for Brazilian textiles, clothing, footwear, citrus products, sugar, etc. To the extent that the FTAA will help these Latin American countries recover from their economic and growth stagnation, it will strongly serve Brazilian and MERCOSUR trade interests.

2.6 Other Free Trade agreements among 2 countries in the Western Hemisphere

There are other free trade agreements that are worthy to mention. There is the Chile MERCOSUR free trade agreement in which the members of MERCOSUR have a free trade arrangement with Chile. The objective of this agreement is to maintain and expand preferential tariff arrangements between Chile and MERCOSUR. This agreement went into effect in October 1996. Chilean imports from the MERCOSUR were $4.5 billion (16% of total imports); MERCOSUR imports from Chile were $1.7 billion (2% of total imports).

Another free trade area is the one formed by Chile and Mexico in September 1991. The objective of this agreement is to promote bilateral trade and investment flows. The other objectives included the elimination of tariffs on 90% of the traded goods, tariffs phased out in synthetic textiles, glass, ceramics, meat, poultry, eggs and some timber products. The tariffs on vehicles should be reduced and there should be harmonization in tax and investment rules. Although this agreement went into effect in 1992, tariffs on nearly all products were gradually reduced and eliminated in 1996. Tariffs on more than 100 products were abolished by 1998. However, there are products that are excluded from tariff cuts such as sugar, tobacco, and petroleum products.

The most recent Free Trade Agreement was signed between the US and Chile (2002). This pact says that tariffs and quotas on all goods should be abolished after the transition period with no exemptions. With this agreement, about 85% of the trade in consumer and industrial goods becomes duty-free when the agreement is signed. The remaining tariffs will be eliminated within four years after the agreement is implemented. The US exports that will gain free access to the Chilean market are: agricultural and construction equipment, auto and auto parts, computers, technology products, medical equipment and paper products. Textiles and apparel will have zero tariffs if they meet the rules of origin agreed in the pact. This agreement is definitely an open door to the conclusion of the FTAA by the specified date. This agreement is a comprehensive one that includes trade in services such as banks, insurance, securities and related firms, open telecommunication market, open investment flows, high level of intellectual property protection, protection against anti-competitive and monopolistic behavior, dispute settlement, etc. In general, the structure of the Chile-US free trade agreement is similar if not equal to the proposed structure of the FTAA. This agreement covers all the relevant areas that need attention and it definitely can be used as a model to start the FTAA agreement.

2.7 Special Problems: Rules of Origin in the FTAA

When countries join a Free Trade Area, there are complex rules of origin that are part of the agreement. In a free trade area, the member countries will have zero tariffs among themselves. However, every country can keep its own tariff rates for non-members. Rules of origins were created to prevent a non-member from exporting a product to a low tariff member and then re-exporting the product to a high tariff member without paying the higher tariff. The rules of origin determine whether goods qualify for preferential tariff treatment. Products that are
wholly made from material produced in the FTA qualify for free trade. Products imported from non-members do not. Products containing some materials imported from non-members may or may not qualify for free trade. This will depend upon the rules of origin. These rules that should be created for the FTAA will probably be based on the existing ones in NAFfA. In any instance, rules of origin represent a set of requirements that are based on three conditions that need to be met in order to have goods exported under free trade area agreement (Murray, 84):

1. The products should be shipped from the FTAA member country to the other FTAA member without intermediate trade or processing.
2. The products that are being traded need to have appropriate documents that certify that they qualify for preferential tariff treatment under the agreement.
3. The exporting country should have made a minimum processing in the products being traded.

For condition 1, there is an exception when the exporting country is land-locked (Bolivia and Paraguay). In these cases, the products can be in kept in third countries only under the condition of being in transit. Although this is an exception to rules of origin, it becomes complex when there is a free trade area agreement. The rules of origin in NAFTA ensures that free trade benefits are given to firms and individuals who produce or manufacture good in North America using local input materials and labor. NAFTA has agreed to implement many uniform customs procedures and regulations in the three countries that are members. This uniformity in the procedures facilitate the exporting process because it saves transaction costs to the small and medium sized companies that exports to the different NAFTA countries. In addition, the documentation pertaining to rules of origin, records keeping and origin verifications are the same for all three NAFTA countries. In the case of the FTAA this can be a potential problem since there are 34 countries involved in the agreement. In addition, there are different languages spoken in the countries: Spanish, English, French, and Portuguese. NAFTA uses a formula in order to calculate the local value content of the good that is being imported to the member country. The exporter or the producer of the good can choose the transaction value method or the net cost method to calculate this content.

The transaction value method is:

\[ RVC = \frac{TV - VNM}{TV} \times 100 \]

where,

- \( RVC \) is the regional value content, expressed as a percentage.
- \( TV \) is the transaction value of the good adjusted to a F.O.B basis
- \( VNM \) is the value of non-originating materials used by the producer in the production of the good.

The net cost method is:

\[ RVC = \frac{NC - VNM}{NC} \times 100 \]

where,

- \( RVC \) is the regional value content, expressed as a percentage;
- \( NC \) is the net cost of the good; and
- \( VNM \) is the value of non-originating (imported) materials used by the producer in the production of the good.

Under NAFTA, a product is said to originate in the free trade area when it grows, is harvested, wholly produced, or substantially transformed in the free trade area. When substantial transformation is the case, then the process causes a product to shift from one tariff classification to another. The term substantially transformed still in controversy sometimes. For products to qualify for free trade there should be a limit on the use of imported inputs when processing the good. This has the objective of having the value of the imported inputs not to exceed a certain percentage of the export value of the final good. Rules of origin should be used for the only purpose of avoiding the “free rider” problem. Generally, rules of origin can be used as a non-tariff barrier to protect domestic industries. With regard to the FTAA rules of origin, there is a need to establish a uniform system for FTAA rules to make the trade process less complex. The system would need clear, transparent and symmetric rules of origin to avoid confusion.

3. My proposed analysis: Brazil and United States

In this analysis, the main focus will be the relationships between The United States and Brazil because of their mutual relevance in the negotiation of the FTAA. It is important to mention that Latin America has a diverse set of economies, social and political structures. The effects of this free trade area will mainly depend on the individual country. However, the purpose of this thesis is to evaluate the effects that this free trade agreement will have on the trade patterns between Brazil and the United States. The effects that will be presented will be from the perspective of consumers, producers, rest of the world and NAFTA. According to the thesis that will be exposed, Brazil and other Latin American countries will be the main beneficiaries of this free trade area. In general, consumers will benefit, domestic producers will lose and the imports from countries outside the free trade area will decrease as well. In this analysis, there will potentially be trade diversion and trade creation between these country players and this will be discussed in the later section of this project. Brazil will be generally mentioned as the exporting country and the United States as the importing country.
3.1 Trade Creation and Trade Diversion

Two of the main incentives produced by a free trade area like the FTAA are: trade creation and trade diversion. We will assume for analytical purposes that the FTAA reduction of tariffs in Brazilian products and US products will become effective immediately after the agreement is signed. This move toward integration and free trade for the hemisphere is occurring against an extraordinary environment in which advanced economies are merging with developing nations. In this case, the United States represents the industrialized nation and Brazil represents the developing nation. Although Brazil is considered an agrarian economy still, it has moved along way in the past decade. Brazil is the dominant member of the MERCOSUR accounting for approximately 70% of the total GDP, about 80% of its population and two thirds of its total trade. (IDB-Intal 2000, Schott). Brazil has contributed to the region outgrowth because of investment in important transportation and telecommunications infrastructure and in energy sources. Brazilian exports are gaining an ever-increasing share of the world markets for manufactured goods. Brazil is the world’s largest producer and exporter of coffee but coffee exports account for only 5% of total exports. Brazil’s largest single trade partner is the United States. The United States imports a wide range of products from Brazil, from orange juice concentrate to automotive parts, shoes, textile, airplanes, etc.

When the FTAA takes effect, there will be trade effects. If we analyze this from the prospect of Brazilian welfare, we can say that the FTAA will eliminate the tariff rates and as a result, the United States will import more from Brazil increasing US consumption, displacing domestic production, displacing imports from the rest of the world and displacing imports from NAFTA. This will be the general impact that applies to all products whether they are final goods or input materials. Other Latin American countries will also share this effect because they will also be part of the FTAA. Of course, it is important to note that it will depend on the type of product that is being analyzed. There may be cases where Brazil will be the only FTAA country that exports a certain product to the US and in this case, the other Latin American countries will not have any shared effect. Trade creation (Figure 1) means that the free trade area will create trade between the exporting country and the importing country that would not have existed otherwise. As a result, the supply of goods shifts to a more efficient producer of the good. In all cases, trade creation will raise welfare. In this base scenario, it is necessary to make a distinction between the effects of the trade creation:

1. Positive impact on US buyers of imports. These could be consumers of final goods and firms using input materials.

3. Negative impact on US imports from NAFTA (Canada, Mexico, Chile)
4. Negative impact on US imports from the rest of the world.

The consumers of the importing country will benefit from the free trade area because of the reduction in the domestic price of both imported good and domestic substitutes, raising consumer surplus. The producers in the importing country will suffer losses as a result of the FTAA. The decrease in the price of their product in the domestic market reduces producer surplus. The price decrease will also cause the decrease in output of existing domestic firms and potentially some firms will shut down. This will have a negative effect on employment and a decrease in profits. As a result of this effect, the domestic producers will have to accommodate to the competition in low cost production. The government of the importing country will suffer a loss of revenue that will be transferred to the exporters of the exporting country. It is important to realize that when the FTAA becomes a reality, many markets and multiple countries will be affected. In order to analyze the aggregate effects of the FTAA, we would need to sum up the effects across markets.

When the FTAA comes into effect, there will also be another effect that is called Trade Diversion (figure 1). The free trade area will divert trade away from a more efficient supplier outside the FTAA towards a less efficient supplier within the FTAA. The effect of trade diversion can be positive or negative for national welfare and that will depend on how much trade creation is born from the FTAA. To illustrate trade diversion, here is a simple example: assume that the United States was importing plastics from Spain at a cost of $3 per feet plus $1 tariff; the import price is $4. The cost for Brazil is $3.5 per feet. When the US and Brazil form the FTAA, the tariff rate imposed in the Brazilian plastic will be zero and the plastic price will be $3.5 per feet. In this case, there will be trade diversion from the plastic supply from Spain to the plastic supply from Brazil.

The welfare effects of trade diversion are negative. If trade creation is larger than the magnitude of trade diversion, then the net welfare will be positive. On the other hand, if trade creation is smaller than the magnitude of trade diversion, then the net welfare will be negative. Generally stating, the larger the difference between the FTAA country and the rest of the world, the more likely that trade diversion will reduce welfare.

Evaluating the effects of trade creation and trade diversion from the perspective of Brazil, it can be concluded that with the introduction of the FTAA, Brazil’s economy will experience what is called a Trade Expansion that benefits Brazilian economy. It is necessary to recognize that there are preferential agreements between the US and Brazil in trade of certain products such as airplanes and coffee. In the case of these products, the FTAA will not have a positive or negative effect on the trade between these two countries. This is also true for other Latin American countries.
that benefit from free access to the US market. In this case, these other Latin American countries (e.g., Mexico, Chile) are already exporting their products duty-free and the introduction in the FTAA will not cause an increase in the volume of trade with the US. It might even hurt them since the US will be able to import from more suppliers without tariffs.

Impact of Free Trade Agreement

3.2 Impact of the FIPAA on Brazil

From the Brazilian perspective, the FTAA could potentially facilitate the access to the US market, lower the costs of inputs and final products, facilitate the transfer of technology and increase the investment flows towards all Latin American countries. The United States has also interests in the FTAA. Among these interests are the stronger economic, political and foreign policy ties with Latin American countries, especially with Brazil. The US is looking forward to the open access to the large Brazilian market of goods, services and capital movements throughout the region. The United States wants also to get the share of the trade that the European Union has conquered in Brazil. The FTAA would put the US in advantageous position vis-à-vis European and Asian companies that are currently doing business in this area. Brazil is an important trade partner of the US. During the 1990's, Brazil implemented market oriented reforms where protectionist policies were replaced by privatization of inefficient government-owned agencies and the liberalization of trade and investment. According to the SECEX (Secretary of External Commerce) of Brazil, in 1999 Brazil's top exports were mainly comprised of natural resource based goods, and manufactured goods only accounted for 27% of Brazil's top ten exports. In the year 2000, the US accounted for 23.9% of Brazilian exports. The European Union accounted for 26.8%, and the rest of Latin America accounted for 23.4%. This indicates that almost 50% of Brazilian exports go to the Americas. With the FTAA, the Brazilian exports will be focused on the US market at the expense of European countries. Brazil is a very attractive trade partner for the US. American companies have been realizing that the opportunities to invest in Brazil are enormous. More than 400 of the Fortune 500 companies currently have operation in Brazil. Brazil has also been identified by the US Department of commerce as one of the ten "strategic partners" of this century.

Brazil imports have been diversifying throughout the last years. Approximately 49% of the imports are raw materials and 27% are finished goods. The imports of durable goods are led by automobiles that are among the main products imported by Brazil. The trade flow between Brazil and the US member countries is very strong and is one of the ten "strategic partners" of the US. Brazil has also been identified by the US Department of commerce as one of the ten "strategic partners" of this century.

3.3 Effective Protection Rate (EPR)

A tariff is a form of protection for the domestic industry. If the tariff is an ad valorem tariff proportional to the value of the imports, then the tariff rate itself represents the amount of protection. If the tariff is a specific amount, then in order to measure the amount of protection, the tariff is divided by the price and the result is the ad valorem equivalent. However, sometimes the effect of a tariff can vary in the different stages of production of a good. Most wealthy countries have an escalated tariff schedule with lower tariffs on raw materials, modest tariffs on inputs materials and higher tariffs on final goods. This escalation causes the effective protection that is highest on final goods, lower on intermediate inputs, and lowest on raw materials. Such escalated structure of tariffs alters the composition of imports favoring the importation of raw materials and discourage...
To illustrate this concept, here is an example:

Suppose the price of a final good is $10 and the price of the input material is $4. Before any tariff the Value Added is $10 - $4 = $6

Suppose the US imposes a tariff (T) on the final good of 20\%, then the final good price is now $12. The Value Added is $12 - $4 = $8 or (($8 - $6)/$6) = 33\%

Suppose that the US imposes a tariff (t) on the input material of 20\%; then the price of the input material is $4.8. The Value Added is $12 - $4.8 = $7.2 or ((7.2 - $6)/$6) = 20\%

Suppose that now the tariff (t) on the input material is 10\%. The price of such would be $4.4. The Value Added is then $12 - $4.4 = $7.6 or (($7.6 - $6)/$6) = 27\%

Suppose that now the tariff (t) on the input material is 40\%. The price of such would be $3.6. The Value Added is then $12 - $3.6 = $8.4 or (($8.4 - $6)/$6) = 7\%

The effective protection rate (EPR) will be greater when the tariff on the final good is greater than the tariff on the input material. When the tariff on the final good is equal to the tariff on the input material, then the effective protection rate will be the same to both.

\[
\text{If } T > t \\
\text{EPR} > T \\
\text{If } T = t \\
\text{EPR} = T = t
\]

3.4 Trade Creation and Trade Diversion for Dutiable Products from Brazil

The data is based on statistics from 2001. The descriptions of the products in this sample are found in the website of the US International Trade Commission www.usitc.gov, under the Harmonized Tariff Schedule of the United States.

\[
\text{TC} = E_M \times M_B \times \Delta t^B \\
\text{TD}_{\text{NAFTA}} = TC \times \frac{M_{\text{total}}}{\text{US}} \times M_{\text{NAFTA}} / M_{\text{total}} \\
\text{TD}_{\text{ROW}} = TC \times \frac{M_{\text{total}}}{\text{US}} \times M_{\text{ROW}} / M_{\text{total}}
\]

The total value of US imports of sample product from Brazil for 2001 was $1,613 millions. There were a total of dutiable products of $4,981. We can conclude that 32\% of US dutiable imports from Brazil come from this group of products. (See Appendix).

\[
\sum TC_B = $288.79 \text{ millions} \\
\frac{207}{288.79} = 72\% \text{ (tariff revenue)} \\
\sum TD = $186.81 \text{ millions} \\
\sum TE_B = $475.60 \text{ millions}
\]

I conclude that this sample of products contain about 32.4\% of all the products that US imports from Brazil. However, these products account for 72\% of the duties collected. In the year 2001, the total tariff revenue collected from Brazilian imports was $287.97 millions and the sample products accounted for $288.79 millions per year. Clearly, this sample represents the most significant group of products that will benefit from the FTAA. The resting 68\% of products that come to the US from Brazil are not representative in dollars volume or they enter the country with preferential treatment already. Based on this group of significant products, I have also calculated trade creation for Brazil, trade diversion for NAFTA and the rest of the world and finally trade expansion for Brazil. When calculating trade creation:

\[
\text{TC} = E_M \times M_B \times \Delta t^B \\
\text{TC} = E_M \times \text{DUTY}^B
\]

The assumption of these calculations is that the elasticity of imports from Brazil (E~) is 1

\[
\text{TC}_B = $288.79 \text{ millions}
\]

Trade diversion for NAFTA and the rest of the world has also been calculated:
they have a competitive advantage. Footwear exports have the potential to grow exponentially after the FTAA. The calculation Brazil will have to take action in the footwear industry, where textiles and apparel will be completely eliminated for China by 2005.

The benefits that Brazil will get with this agreement will be shared with American countries because the FTAA will include them and Brazil's potential in this industry. According to the Uruguay Round and the Multifibre Arrangement (MFA), the quotas on textiles but not as much as estimated. As a result of this, China has a competitive advantage in this industry and the imports that come to the US. Trade diversion for NAFTA in orange juice is $7.4 millions per year. However there may be an underestimation since Brazil can potentially take over all of the imports of orange juice in the US. Trade diversion as a total for NAFTA is $7.9821 millions and most of it is taken by orange juice. Trade diversion for the rest of the world is $1134.5 millions. Brazil will probably pick up on this amount of trade from other countries like in Europe. US producers will most likely not benefit from this because of cheaper imports to the US. US consumers will have access to a broader range of products at cheaper prices as a result of the abolition of import tariffs for Latin America. Trade expansion for Brazil has been calculated to be $459.58 millions that include products that were not taken into account in the sample (the remaining 28%).

By multiplying the sum of total trade expansion calculated in the sample by one-third, I am accounting for other products that are not in this sample and also for possible errors. When I calculate trade diversion, I excluded the data for other Latin American countries because the FTAA will include them and they will get the same import treatment (0 tariffs) as Brazil. The benefits that Brazil will get with this agreement will be shared with other Latin American countries.

It is important to acknowledge that Brazil exports textiles and apparel to the US and they have a large potential of increasing these exports. However, there may be another factor affecting Brazil's potential in this industry. According to the Uruguay Round and the Multifibre Arrangement (MFA), the quotas on textiles and apparel will be completely eliminated for China by 2005. China has a competitive advantage in this industry and Chinese exports of textiles to the US will grow dramatically when the quotas completely phase out in 2005. Brazil will be able to export textiles but not as much as estimated. As a result of this, Brazil will have to take action in the footwear industry, where they have a competitive advantage. Footwear exports have the potential to grow exponentially after the FTAA. The calculation of trade creation for Brazilian footwear is $114.3. This number is underestimated because Brazil's capacity in manufacturing shoes. This accounts for ($114.26/$207.45) = 55% of the total trade creation calculated for Brazil.

3.5 Special Cases

3.5a Sugar

Sugar is an important product for both Brazil and the US. However, the US has always been very protective with the imports of sugar. Sugar has suffered quantitative restrictions, duties, and fee on imports. Brazil is a world leader producer of sugar. They have a huge capacity of production and without discussion sugar has always been one of the major export products for Brazil. This is the reason why Brazilian sugar is taking a considerable importance in the negotiations of the FTAA. As one of the largest sugar producers, Brazil has considerable influence over the international sugar market. Brazil exports approximately 100,000 to 200,000 metric tons to the US every year but it has a larger role in the global market. In the year 2000, Brazil exported about 8 million metric tons of raw sugar. Currently, the world price of sugar is about 08.5 per pound while the US price is about 022.5 per pound. Brazil is currently the second largest quota holder to the US market according to the Office of the US Trade Representative. Brazil's quota is 152,691 metric tons for 2003. The total quota that the US has on sugar is 1,117,195 metric tons, which is the minimum level under the Uruguay Round Agreement in 1995. This quota is divided among 40 countries. Based on this information, Brazil could take over the imports of sugar to the US after the quotas are phased out. If the FTAA becomes a true free area, Brazil could become the main exporter of sugar to the US. The other main exporters of sugar are: Australia (87,402), Dominican Republic (185,335), Guatemala (50,546) and Philippines (142,160). Brazil will be able to compete with the Caribbean islands that are large producers of sugar too. The reason to say this is because Brazil has a territory advantage over the islands. Brazil's capacity is much larger than any of these small islands. Another potential competitor that we need to consider in a future is Cuba. Currently, Cuba is excluded from this free trade area. However, when Fidel Castro is not in power anymore, there is the possibility that the trade relationships would be built again with the US and this could make Cuba a potential exporter of sugar. A factor to consider in this industry and the imports that come to the US is that the other competitors are the Caribbean Islands. However, these economies are probably producing at their maximum capacity and currently they receive a price four times larger than the world price when they export to the US. When the FTAA is signed, these islands may even be hurt because of the fact that the US will pay world prices for sugar imports.

3.5b Orange Juice

Citrus products are an important negotiation area in the FTAA. In the year 2001, the US imported $75 millions in orange
I am not estimating any investment effects, any Brazilian imports from the US or any Brazilian trade with other Latin American countries. These calculations would have to take more complex processes. However, it is well known that FTAs create incentives for foreign direct investment (FDI) as well as domestic investment. There are two incentives that are important:

1. There are several firms in the US that will have an incentive to invest in productive capacity in Brazil designed to increase production to serve the Brazilian and US markets.

2. Countries that are non-members of the FTAA (EU, Japan, etc) will have similar incentives to invest in productive capacity in Brazil. Since exports from their home countries do not have the preferential treatment that the FTAA countries will face, these outside countries will have an incentive to invest in FTAA members. These effects are complex to estimate.

4. Conclusion

After analyzing the process in which the FTAA is taking place and the potential effects that this will have in the trade between Brazil and the US, my conclusion is that the FTAA will have potential gains for countries that are in the process of development such as Brazil. The poorer countries are the ones who will benefit the most because of the innovative technology and the open access to advanced markets like the US. Brazil in this case, is taking the majority of the benefits from the FTAA. Brazilian producers will benefit because they will be able to export more to the US, especially in the cases of sugar, orange juice and footwear. US producers will potentially suffer losses because of the introduction of more competitive raw materials, intermediate inputs and final goods. The FTAA will undo the escalation of tariff schedules that the US have on finals goods and raw materials. This will give the industries of final goods in Brazil the incentive to export higher volumes of goods to the US. US producers will have to adjust to competitive prices and quality such as orange juice. The cost of production and labor in the US are higher than in other countries and this is a controversy because of labor standards that should be implemented in all countries. However, the fact is that the lowest cost is in Brazil and this will force US producers to catch up in the competitive way of producing output. Because of the FTAA, there will be trade diversion for goods coming from NAFTA and the rest of the world. Based on the findings from the sample of products used to calculate trade creation, trade diversion and trade expansion, the conclusion is that trade diversion will mostly hurt the rest of the countries that are not NAFTA or FTAA members. The majority of the products that Brazil exports to the US are mainly exported by non-NAFTA members. The numbers that came out of the calculations may be underestimated in cases such as sugar, orange juice, tobacco and footwear. On the other hand, calculations for the textile and apparel industry are overestimated because of the elimination of quotas that will take place in 2005 for China. Brazil will have a trade expansion of $459.58 millions and this number is below what may happen when the free trade area comes into effect.

All these benefits will also be shared with other FTAA beneficiaries that are active exporters to the US. It is important to also consider the probability that after the FTAA is implemented, the political, social and economic relations among these countries will become closer and easier. This is clearly a sign of the way in which the world is going, Globalization. However, these are only estimates based on past data on trade statistics. It could be possible that there may be conflicts arising in the process of accomplishing the FTAA because of the complexity issue of such negotiation. There are 34 countries involved with different cultures, languages, currencies, technologies, political structures, etc and this adds to the complexity of forming a free trade area. This will take a long time until the world can see results but let's compare to the European Union. They took decades to form a true union and they had to struggle over the years. It is a union that still in process of fortifying but it sure has the potential of becoming a world power in the international environment. The FTAA can be a possible union but a long process of changes and reorganization has to come along with it.

References:


"NAFTA for the Americas: Q & A on the FTAA (Free Trade Agreement of the Americas)." *Multinational Monitor* Apr. 2001.


www.ftaa-ala.org
www.mac.doc.gov
www.ustr.gov

Faculty Comment:

In his letter to the Inquiry Publication Board, Ms. Cadario’s faculty mentor Professor Tracy Murray wrote:

I have been working with Ms. Cadario for several months and I am extremely impressed with her dedication and originality. She is from Bolivia and majoring in international business, so her interest in an FTAA seems natural. Nevertheless, she chose the topic without input from me. Further, she recognized that estimating the trade effects for all countries in the hemisphere would be a bigger task than she could accomplish under the time constraints of an honors thesis. She decided to limit her inquiry to Brazil, which is the biggest economy in the region and more populous than the rest of South America combined. She carefully analyzed the likely trade effects on her own. At this point, I suggested the estimation techniques that might be useful. She now has the required data on trade flows between Brazil and the US; her analysis will be extremely detailed and based on actual trade flows of those products of export interest to Brazil.

She also recognized from the beginning that the effects of an FTAA would depend upon the pre-FTAA trading environment. Currently, the US has free trade agreements with Canada and Mexico (NAFTA) and four other countries. The US has several preferential trading arrangements that already provide duty-free access to the US market. Thus, the FTAA is an additional trade policy initiative, which adds to the complexity of estimating the likely trade effects. In her preliminary reading she reviewed the rules and regulations of NAFTA that are likely to serve as a model for the FTAA. She noted that the NAFTA rules of origin add additional complexities that must be taking into consideration. Finally, upon identifying the major products of export interest she noted some special cases, namely sugar, orange juice, and textiles and apparel that again must be studied in some additional detail.

The rigor of her analysis and use of research methodologies would be impressive for a graduate thesis. It is doubly so as an undergraduate honors thesis.
THE IMPACT OF INTRASTATE VARIATION IN HIGHER EDUCATION FUNDING ON INTRASTATE RESEARCH AND DEVELOPMENT EXPENDITURES

By Stephanie Nicole Gosnell
Department of Finance

Faculty Mentor: Carol Reeves
Department of Finance

Abstract:

The paper examines the question "Does specialization in higher education result in improved economic outcomes for a state as measured by increased research and development (R&D) in the state?" A fixed effects model is employed to estimate how the variation in state funding per pupil across institutions of higher education (a measure of specialization) impacts R&D funding in the state. Expenditure per pupil data from the Integrated Postsecondary Education Data System (IPEDS) from 1992-2000 for the approximately 600 public, 4-year institutions in the U.S., is used to capture the variation in state funding in institutions of higher education. The results of this study indicate that an increase in the proportion of a state’s funding appropriated to higher education leads to a statistically significant increase in R&D expenditures in that state. The policy implication of this finding is a greater proportional investment in higher education implies a significant return on investment. The study also indicates that an increase in the variation of state expenditures per pupil leads to a positive, but not statistically significant, increase in R&D expenditures in that state. The data suggest, albeit weakly, that specialization in higher education funding leads to improved economic outcomes for the state as measured by R&D expenditures.

"There was that law of life, so cruel and so just, that one must grow or else pay more for remaining the same." - Norman Mailer, The Deer Park

Introduction:

The role of education is to facilitate the realization of the potential of each citizen, the economy, and society. Higher education is specifically undertaken to create opportunities for moving beyond our current state. However, investments in creating such opportunities come at the expense of other programs that require public funds. Therefore, the efficiency of education funding is a particularly crucial matter due to the stagnant, if not decreasing, resources available for higher education. With the strain on these funding resources, it is imperative to allocate them in the most organizationally efficient and effective way. The introduction of new graduate degree programs and the duplication of degree programs which is occurring within Arkansas and throughout the nation raises questions regarding the logic behind the expansion of expensive graduate programs which must come at the expense of alternate uses for the funds including targeted funding which encourages research activity at research institutions. The potentially inefficient allocation of state funds is one factor affecting the current amount of Research and Development (R&D) dollars available to research universities. It is necessary to determine whether those responsible for dispersal of funds are being responsible stewards of the taxpayers’ money. To responsibly allocate funds, legislators must be aware of structural efficiency issues to correct any structure-based problems in allocations of educational funding. They need to have relevant information regarding the impact their decisions can and do have upon economic factors such as education. More importantly, they must understand that failure to maximize educational outcomes implies the state’s economy is operating at less than its potential.

Literature Review:

The role of knowledge in economic growth is an increasingly popular topic. Economists are continually adapting traditional growth models to account for science and technology applications stemming from higher education. The general link between specialization, increased productivity, and increased economic outcomes has been established for several centuries. Adam Smith’s

The Wealth of Nations, which was published in 1776, was one of the first works to establish this link and opened the door for future research. The work opened with a description of the
manufacturing of pins and describes how specialization of labor increases the productivity of the workers as a whole (Landry, 1997). In the production process, he detailed that those with strengths in an area should take on the responsibility for that area with the phrase, "as one man draws out the wire, another straights it, a third cuts it" (Fajardo-Acosta, 2003). His work had two main themes. The first involved how increasing the division of labor increased the productivity of labor. He emphasized specialization as a key factor in this increase and pointed out a desire for a higher standard of living as the motivation behind it.

The second theme related to the limits of division of labor depending upon the size of the market. He stressed that large markets are essential to the division of labor and to high productivity (King, 2003). Smith discussed the link between higher productivity and higher income, and then connected this to increased demand and larger markets. He discussed the propensity to exchange, which leads to division of labor, thereby increasing productivity (Kilcullen, 1996).

Many may argue that technology is responsible for advancement in productivity. Adam Smith argued that the division of labor enabled technology to develop and progress, and therefore, specialization of labor is the key to material well-being (Kilcullen, 1996). Our standard of living is affected by three variables: productivity of labor, the division or specialization of labor, and the size of the market (Kilcullen, 1996). Smith expanded upon the policy of Europe and how irresponsible allocations of public funds for education only led to economic downturn (Kilcullen, 1996). On the division of labor Adam Smith specifically stated:

"Those ten persons, therefore, could make among them upwards of forty-eight thousand pins in a day. Each person, therefore, making a tenth part of forty-eight thousand pins, might be considered as making four thousand eight hundred pins in a day. But if they had all wrought separately and independently, and without any of them having been educated to this peculiar business, they certainly could not each of them have made twenty, perhaps not one pin in a day; that is, certainly, not the two hundred and fortieth, perhaps not the four thousand eight hundredth part of what they are at present capable of performing, in consequence of a proper division and combination of their different operations (The Library of Economics and Liberty, 2003: 5)."

While it may be one of his more quoted phrases, Smith used the "invisible hand" analogy only two times in his publications.

"every individual necessarily labours to render the annual revenue of the society as great as he can. He generally, indeed, neither intends to promote the public interest, nor knows how much he is promoting it. By preferring the support of domestic to that of foreign industry, he intends only his own security; and by directing that industry in such a manner as its produce may be of the greatest value, he intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention (2003, 5)."

These comments are applicable to the higher education system in the United States. Education and educators are the hand guiding our youth to prosperity, and in doing so, stimulating the economy through increased productivity and technological advancement. Smith's work has been long established, and the fact that specialization generally leads to better outcomes is proven; however, the link between specialization in higher education and improved economic outcomes has not been established.

Another noteworthy economist, David Ricardo used the example below to illustrate the importance of specializing in the most efficient ways according to strengths and to avoid duplication of work if it is not necessary or the most efficient means:

"To produce the wine in Portugal, might require only the labour of 80 men for one year, and to produce the cloth in the same country, might require the labour of 90 men for the same time. It would therefore be advantageous for her to export wine in exchange for cloth. This exchange might even take place, notwithstanding that the commodity imported by Portugal could be produced there with less labour than in England. Though she could make the cloth with the labour of 90 men, she would import it from a country where it required the labour of 100 men to produce it, because it would be advantageous to her rather to employ her capital in the production of wine, for which she would obtain more cloth from England, than she could produce by diverting a portion of her capital from the cultivation of vines to the manufacture of cloth (2003: 3)."

In the 21st century, program duplication in university systems within states is one example of a domestic problem that leads to financial inefficiencies. A study commissioned by the Arizona Board of Regents in 1988 studied the internal and external needs for program duplication, the need to avoid duplication in the absence of a need for duplication, why avoidance was a more efficient system than elimination of programs once they were in place, and how university structures and systems do and could review programs before they are implemented. The study emphasized the need for accountability for resources allocated to public universities through continual review of the procedures used to both begin new programs and sustain the existing programs. Other studies pertaining to program duplication, especially those in Montana and Colorado were reviewed by Arizona (Macvicar, 1988).

Another study by Owen Cargol in 1983 stressed the importance of limiting program duplication and referred to the topic as "a bugaboo in discussions of higher education" (Cargol, 1983:2). Cargol discussed the decline in government support for higher education that caused budget cuts that may result in selective or sweeping cuts. He discussed the varying levels of
program initiation policy by state. The report details the current policies for program review and creation. Fifteen states currently practice a "review and recommend only" responsibility for existing programs. State boards or commissions are given "review and approval power" in thirty-one states (Cargol, 1983: 3). Cargol contended that this level of oversight was not enough to push our institutions to become more efficient on a statewide basis.

The connection between improved economic outcomes and university research has been established on an international level. One of the more recent works is a study by Fernand Martin in 1998. His examination of Canadian university research revealed that university research is a powerful stimulus for economic development, producing measurable increases in GDP and employment. According to his study, university research accounted for one percent of Canada's GDP and more than .05 percent of all jobs. He also found that university research had a "profound effect on the underlying productivity of the economy" (Martin, 1998: 2). His report revealed that university research equipped students with the ability to generate new ideas and that companies benefited from this research by hiring graduates with knowledge and research skills. "The total dynamic impact of university research is approximately $15.5 billion each year, which is equivalent to about 150,000 to 200,000 jobs" (Martin, 1998: 2). Martin used Total Factor Productivity (TFP), the economic growth that results from increases in the efficiency and productivity of labor and capital, to quantify productivity growth. While his work provided a convincing link establishing the importance of university research to economic growth, it did not suggest specific methods for comparing specific institutions and making university allocations more efficient, nor did it examine the impact of university research in the United States.

Several recent studies discuss growth models in relation to R&D in the United States. One such study by Charles Jones suggested, "growth is generated endogenously through R&D and growth in the economy is tied directly to growth in productivity, which in turn depends on the discovery of new designs through R&D" (Jones, 1995: 759). Another study seeking solutions to the European Union's attempt to close the gap with the US in income per capita states that investment in human capital, through education, and R&D are "essential for high productivity in all industries", especially high technology industries (Corley, et al.; 2002). Two studies by Paul Romer published in 1986 and 1990 in the Journal of Political Economy discussed previous growth models and the theoretical gaps left by them due to dependence on exogenously specified population growth and its relationship to per capita income. The gap was widened by "the loose treatment of specialization as a form of increasing returns with external effects" (Romer, 1986: 1034). He attempted to fill the gap by providing a model with both increasing marginal productivity of knowledge and decreasing marginal productivity of physical capital (Romer, 1986). The second study by Romer created a model where growth is driven by technological change arising from "intentional investment decisions made by profit-maximizing agents" (Romer, 1990: S71). It concluded that human capital determines the growth rate of the economy and subsidizing the accumulation of total human capital is positive (Romer, 1990).

In September 2002, the Milken Institute's State Technology and Science Index Research Report for the U.S. provided a current link between increasing the knowledge base of a state's population through university degree programs and research and increased economic outcomes. It detailed the "intangible economy", those factors that contribute to economic outcomes, and the importance of research in developing the economic progress within a state. Stress was placed upon the importance of higher education for direct research, providing a knowledge-based workforce, attracting industry, and the creation of technology clusters. These technology clusters have a tendency to develop and remain in regions with existing research and development operations, such as institutions for higher education. (DeVol, 2002). A key finding in the report states "those states with vibrant technology clusters will experience superior economic growth" (DeVol, 2002: 5). The report also finds that human capital is driven by the ability to attract and leverage science and technology assets (DeVol, 2002). The efficiency in university-based research and development, however, limited or expanded R&D's ability to be a driver of economic development. The Milken Report authors argued that R&D dollars must be spent wisely within the states. It used the Academic R&D dollars per capita to illustrate the importance of university research, and outlined a connection between allocations to universities based upon strength and competence of the university systems. It addressed the importance of the percentage of a state population with a Ph.D. due to the need for advanced researchers within a state to enable advancement in industry and further the knowledge and understanding of the surrounding population. While the Milken Report illustrated the importance of science and technology across the states and their impact on economic conditions, it did not pinpoint specific methods that would assist in the appropriate allocation of resources.

Research Methodology:

In this study, specialization in educational funding is measured through the variation of state appropriations to institutions across a state. Economic theory would predict that specialization in higher education would lead to improved outcomes in a variety of directly measurable outcomes. Such direct outcomes include number of undergraduate, graduate, and professional degrees, reputation, R&D dollars, and journal article publications. The quantity of R&D dollars is a proxy for current and future economic vitality. Indirect outcomes of specialization include migration of industry and corporations to areas near institutions to strengthen their employee quality, an increase in per capita income from higher quality degrees resulting in better...
jobs, and an increase in the number and qualification of jobs within a state. Limits on data resources, however, prevent the measurement of several of the direct outcomes, so the measure of R&D dollars within the state provides a proxy for an index of possible outcomes and the variation in higher education funding within a state serves as the measure of specialization. Higher education funding has a direct impact on R&D expenditures in a state and the economic impact resulting from R&D dollars within that state. While program duplication is a good variable to address efficiency issues with regard to state allocations, it does not allow for differentiation between programs on a qualitative basis and is difficult to measure due to data inconsistencies. No study has succeeded, due to measurability, qualitative factors, and data constraints, in providing a direct link between program duplication and a change in educational or economic outcomes.

Research Question:

The research question for this study is “Does specialization in higher education funding result in improved economic outcomes for a state?” Specialization in higher education funding is measured by the differentiation in state appropriations per student a state makes in its funding process between institutions. The improved economic outcomes are measured by the change in R&D expenditures within a state. Factors influencing this outcome, discussed below, are taken into account in the study. State appropriations for higher education are a key factor in determining R&D’s relationship to economic development because they show how much money is allocated to each institution within a state to operate their university systems and the priority level of education within the state, as opposed to other funding categories. Shifts in appropriations to institutions give insight into the focus of the legislature and in state spending patterns. By efficiently funding higher education systems, the capital to maintain and improve academic and athletic programs that attract students to higher education, which in turn provides more capital to research and development projects, is readily available (DeVol, 2002). The lag in the economic value and long-term results of university research has been proven, but there is no indication that there is no short-term payoff to university research.

Hypothesis:

Total Research and Development (R&D) expenditures in state \(i\) in time \(t\) at institutions of higher education is a function of Gross Domestic Product at \(t\) (to account for economic lag), variation in educational expenditures among institutions per capita in state \(i\), state expenditures in higher education as a percent of total state funding, and the number of scientists and engineers in the state. The primary question of interest is, as state \(i\) specializes its educational expenditures, does R&D funding experienced in the state change? Mathematically, this relationship can be expressed with the following equation; using a fixed effects model to explain the level of research and development (R&D) expenditures across states for the period 1992 through 2000.

\[
RD_t^i = \alpha_i + \beta_1 RD_{t-1}^i + \beta_2 \text{HEDEXP}_t^i + \beta_3 \text{PhD}_t^i + \beta_4 \text{PhD}_{t-1}^i + \beta_5 \ln(Va_t^i) + \epsilon_t^i \quad [1]
\]

Variables and Data Collection Procedures:

\(RD_t^i\) is the dollar amount of R&D expenditures (in millions of dollars) in state \(i\) in year \(t\), and \(RD_{t-1}^i\) is the lagged dollar amount of R&D in state \(i\). The R&D data come from Table B-29 of the 1992- through 2000-edition of the National Science Foundation publication entitled “Academic Research and Development Expenditures.” The table provides the total dollar amount of R&D expenditures at each institution of higher education in each of the fifty states. The table also provides the sources of the R&D expenditures at each institution, e.g., federal government, state and local governments, industry, and institutional funds. (See Appendix A-1.)

The R&D values for 1991 come from Table B-23, “R&D expenditures at doctorate-granting institutions, by geographic division and state: fiscal years 1985-92,” of the same NSF publication. All dollar amounts were adjusted for inflation to 1996 constant dollars using the Gross Domestic Product (GDP) implicit price deflator from the U.S. Commerce Department, Bureau of Economic Analysis (BEA). For the purposes of this analysis, the total amount of R&D expenditures in each state is used.

For the period 1992 through 1995, Table B-29 displays data for R&D expenditures at doctorate-granting institutions. For the remainder of the sample period, the data from the table displays R&D expenditures at universities and colleges (both doctorate-granting and non-doctorate-granting institutions). Technically, this is comparing apples (the R&D data for 1992-95) to oranges (the R&D data for 1996-2000); however, the proportion of R&D expenditures at universities and colleges that is attributed to non-doctorate granting institutions is very small and would therefore have a negligible impact on the results of our estimation. The importance of the R&D expenditures lies in the impact of the expenditures in driving technological advancement, and hence, driving economic growth. The level of expenditures is directly impacted by the variables below.

\(\text{HEDEXP}_t^i\) is the proportion of state expenditures going to higher education in state \(i\) in year \(t\). The data come from the annual National Association of State Budget Officers’ (NASBO) State Expenditure Reports 1992-2000 (NASBO, 1992-2000). Some states had missing values for this variable; as such various state budget agencies were contacted (see appendix).
If our estimates are incorrect, the parameter estimate for this variable will be biased either upwards or downwards. The variable is important because the percentages allocated to higher education give an indication of the priority level of the category within the state. This ratio should be positively correlated with R&D expenditures within a state because as a greater percentage of the state’s funds are allocated to institutions of higher education, more funds are available for R&D at the institutions. In turn, this increases the skill level of students that will become the labor force of the state. The data points for the following states and years were filled with the averages from the remaining years: Mississippi (1992), Alaska (1996), Nevada (1992-1994, 1996-1998), and Wyoming (2000). (See Appendix A-2.)

\( \text{PhD}_i \) is the number of doctoral scientists and engineers in state \( i \) in year \( t \), and \( \text{PhD}_{i,t} \) is the lagged number of doctoral scientists and engineers in state \( i \). The data come from Table 25, “Employed doctoral scientists and engineers, by geographic location and broad occupation,” from the NSF publication \textit{Characteristics of Doctoral Scientists and Engineers in the United States} for the years 1993, 1995, and 1997. The 1999 data come from Table 25, “Employed doctoral scientists and engineers, by employer location and broad field of doctorate,” from the 1999-2000 edition of the same publication. A constant growth rate method was used to estimate values for 1991, 1992, 1994, 1996, 1998, and 2000. If the estimates are not correct, the parameter estimates will be biased either upwards or downwards. These advanced degrees are an indication of the state labor force’s knowledge base, skill level, and sophistication. States with high levels of Ph.D. degree holders have quality research and development centers (DeVol, 2002: 82). This variable will be positively correlated to R&D expenditures because as the number of researchers within a state increases, a greater demand for R&D dollars within the state will result. (See Appendix A-3.)

\( \ln(Var) \) is the natural logarithm of the sample variance of state expenditures per pupil (adjusted for inflation to 1996 constant dollars using the GDP implicit price deflator) across four-year institutions of higher education in state \( i \) in year \( t \). The data come from the Integrated Postsecondary Education Data System (IPEDS) from the National Center for Education Statistics of the U.S. Department of Education. The state appropriations per student were derived by dividing the total state appropriations for institution \( x \) by the total enrollment at the institution for a given year. The variance within the state appropriations was found using the state appropriations per student. The variation among those allocations reflects the degree of specialization in higher education funding within a state. (See Appendix A-4.)

There are three major issues with the state expenditures per pupil data. First, not all schools existed for the entire sample period, and other schools had missing values for one or more years. To fill in the missing values, various state budget agencies and institutions were contacted. Not all phone calls were returned; consequently, for each school that had only one year of data missing, that year’s value was estimated by taking the average of state expenditures per pupil at that institution. If the estimates are not correct, the parameter estimates will be biased either upwards or downwards.

Because of the relatively short length of the sample period, schools that had more than one missing year of data were dropped from the data set. Taking the average of the remaining values as the estimate of the missing values would be unreasonably restricting the values of expenditures per pupil. Doing so would also have the potential of making our estimates of the sample variance less robust. Therefore, these observations were dropped from the data set. The following institutions were removed from the data set due to missing values for more than one year: Arizona State East Campus, Arizona State West Campus, California State University Monterey Bay, California State University Channel Islands, San Diego State University, Southern University Law Center, Benjamin Franklin Institute of Technology, University of Maryland University College, Truman Medical Center for Nurse Anesthesia, Nevada State College at Henderson, University of New Hampshire Manchester, Rutgers University Camden, Rutgers University New Brunswick, Rutgers University Newark, Roger’s University, OGI School of Science and Engineering at OHSU, University of Pittsburgh Bradford, University of Pittsburgh Greensburg, University of Pittsburgh Johnston, A&M University System Health Science Center, University of Texas Anderson Cancer Center, Education Service Center Region 2, Washington State University Spokane, Washington State University Vancouver, and Washington State University Tri-Cities.

Schools that did not exist for the entire period were omitted when calculating the sample variance. Visual inspection of the data suggests that when a new school came into existence in a state, per pupil expenditures at other institutions within the state did not change significantly. The number of schools that were omitted from the data set, for either having more than one missing year of data or for not existing for the entire sample period, totaled 25 (from a population of around 600).

The second major issue with these data is that medical schools were counted as four-year institutions. Since expenditures per capita at medical schools tend to be quite large, the existence of these institutions tend to inflate the variance of state expenditures per pupil, suggesting a greater degree of specialization in a state’s higher education system than may actually exist. This would tend to decrease the reliability of the parameter estimate for this particular variable. However, because the R&D expenditure data to medical schools were not explicitly given in the NSF data, it was necessary to keep the medical schools in the data set to compute the sample variance of state expenditures per pupil.
Third, the University of Wyoming was the only university in the state of Wyoming to receive state funding for the sample period. Consequently, the variation in state expenditures per pupil across institutions of higher education in this state is zero for the entire sample period. The natural logarithm of zero is, of course, negative infinity. Therefore, in our estimation of equation [1], the state of Wyoming was omitted. However, this should not have a major impact on the results of our estimation.

Results:

Estimating equation [1] using ordinary least squares gives us the following estimates, shown in Table 1. (See Appendix A-5 for complete results.)

The model implies that, all else equal, a $1 million increase in R&D in year \( t-1 \) will lead to, on average, a $986,000 increase in R&D in year \( t \). Given that many R&D projects span more than one year in duration, this finding is as expected. This variable is significant at the 0.01 level of significance.

Table 1: OLS Estimates of Equation [1]

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R_D_{t-1} )</td>
<td>0.986</td>
<td>0.0293</td>
<td>33.68</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>( HEDEXP_t )</td>
<td>1.215</td>
<td>0.620</td>
<td>1.96</td>
<td>0.05</td>
</tr>
<tr>
<td>( PhD_t )</td>
<td>0.00319</td>
<td>0.00167</td>
<td>1.92</td>
<td>0.06</td>
</tr>
<tr>
<td>( PhD_{t-1} )</td>
<td>0.00787</td>
<td>0.00228</td>
<td>3.45</td>
<td>0.0006</td>
</tr>
<tr>
<td>( \ln(Va_r) )</td>
<td>4.841</td>
<td>3.787</td>
<td>1.28</td>
<td>0.20</td>
</tr>
<tr>
<td>F-value</td>
<td>9016.57</td>
<td></td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Root MSE</td>
<td>21.73</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

According to the model, all else equal, a one-person increase in the number of employed doctoral scientists and engineers in a state in year \( t \) will lead to, on average, a $1,670 increase in R&D expenditures in that state in the same year. This variable is significant at the 0.05 level of significance. The sign of this variable is as expected; the more employed doctoral scientists and engineers in a state increases the likelihood that there are more doctoral scientists and engineers engaged in R&D activities in a state. The number of doctoral scientists and engineers engaged in R&D activities should be positively correlated with the level of R&D expenditures in a state in a particular year.

According to the model, all else equal, a one-person increase in the number of employed doctoral scientists and engineers in a state in year \( t-1 \) will lead to, on average, a $7,870 increase in R&D expenditures in that state in the same year. This variable is significant at the 0.01 level of significance. This result suggests that doctoral scientists and engineers, by their presence, help attract R&D expenditures in a state in the next year, i.e., funding follows the scientists and engineers.

According to the model, all else equal, a one percent increase in the sample variance of state expenditures per pupil in a state in year \( t \) will lead to, on average, a $4.84 million increase in R&D expenditures in that state in year \( t \). Contrary to our a priori beliefs, this variable is not found to be significant at any tolerable level of significance.

There are three possibilities in explaining why this variable is not statistically significant. First, it may be due to medical school's skewing the sample variance and thus overstating the degree of specialization of the system of higher education within a state. Second, the coefficient may not be statistically significant due to the omission of relevant covariates. Third, a longer and complete time series, i.e., no estimated values for state expenditures per pupil or for the proportion of state expenditures going to higher education, would increase the validity and robustness of the parameter estimates. All three of these need to be addressed in future research.

Overall, though, the model appears to be relatively "good." We note that the F-statistic for the model is highly significant. Moreover, an F-test rejects the hypothesis that the fixed effects are jointly zero at the 0.01 level of significance. Lastly, a Durbin-
Watson statistic of 1.893 implies no positive autocorrelation in the error term, and visual inspection of the residuals suggests that the error term follows a white noise process.

Discussion:

The policy implications arising from the results reached in this study relate to the areas of the higher education funding structure and process, state legislative appropriations, and educational attainment and funding in general. Due to the finding that an increase of one percent in funding for higher education from the state budget will result in a $1.2 million increase in R&D funds within a state, legislators need to consider the positive impact of increased R&D funds upon a state's economy, which has been shown in a variety of studies (e.g. the Milken Report), and therefore, consider allocating a larger percentage of state funding to higher education. While this is not an original finding and was therefore expected, it reinforces the positive relationship between higher education funding and R&D. The policy implication of the finding that an increase in employed doctoral scientists and engineers within a state is positively correlated to increased R&D dollars is that states should pursue policies to increase the level of educational attainment of their populations in order to increase the level of R&D expenditures within the state, and by extension, the level of economic growth within the state. Retention of these doctoral scientists and engineers within the state should be a goal as well. Similar results have been reached in other studies, such as the Milken Institute's Science and Technology Research Report, and were therefore expected. The finding that a $1 million increase in R&D in a given year will lead to, on average, a $986,000 increase in R&D in the next year implies that those responsible for allocations of R&D funding should consider the impact their decisions on allocations will have in the future. While the variation in appropriations to institutions within a state was not statistically significant, it did have the positive sign on the coefficient that was expected. If the data had been available for a longer time series and had excluded skewing factors, the variation may have proven to have a major impact upon R&D. The model itself has also been proven to be more accurate than other models explored.

Need for Further Research:

Data limitations and time prevented the length and completeness of the time series and number of institutions that were possible to include in the study. Further research based on a superior dataset would allow for a broader scope, and therefore, provide a more far-reaching and dependable set of conclusions. Any future research will also need to account for private institutions and the possibility that they may have an impact upon the R&D dollars being allocated to public institutions. States with private institutions that receive a substantial proportion of either national or state R&D dollars would most likely have an impact upon the public institutions within that state. The reason for this consideration would be to account for the either complementary or substitution relationship between the private and public institutions. The former implies that a public institution is more capable of generating R&D dollars because of the private institutions' strength, which results in a synergistic relationship. The latter implies that the state may withhold R&D dollars from public institutions because the private institutions are filling the R&D role for the state. This factor should be considered due to the fact that several private institutions dominate their state in receiving R&D dollars within that state and nation. In future research, a longer time series may result in more accurate results. Any future work would also need to consider omitted variables. Finally, accounting for the impact certain medical schools had upon the funding variation within the states is another issue that needs to be addressed. This factor may be the main reason that the variation variable was not statistically significant.

End Notes:

1 The logarithm, instead of levels, was chosen for interpretive purposes.
2 We report the fixed-effects intercepts in the appendix.

Editor's note:

Ms. Gosnell's paper has both an extensive bibliography and appendix. Space limitations preclude the publication of these items in the journal. However, her paper, complete with bibliography and appendix can be found on the Inquiry website.
ORGANIZING MADNESS:
PSYCHIATRIC NOSOLOGY IN HISTORICAL PERSPECTIVE

By Eric D. Jackson
Department of Psychology
Faculty Mentor: Dr. David Schroeder
Department of Psychology

Abstract:

This paper traces the history of psychiatric nosology in the US from its origins in the early 19th century through the most recent revision of the standardized classification, DSM-IV TR. The evolution of nosology is found to be shaped not only by advances in knowledge, but also by socio-historic and professional trends. The initial impetus for systematic classification came from outside the mental health profession, but later revisions reflected intraprofessional struggles and experiences. Future revisions will almost certainly be prone to these same intra- and extraprofessional influences, and may see a dramatic shift away from symptomatology and towards an etiological focus.

Editor's note: Space limitations precluded publication of the entire thesis. The complete paper, including the following omitted sections can be found on the Inquiry website: Early Classification Systems and the Rise of American Psychiatry, Kraepelin and the Movement to Classify, Creating a Standardized Nosology, the Experiences of World War II, DSM-III and the Next Nosological Revolution, and DSM-IV: A Significant Change.

Introduction

In a back page essay that appeared in Time late in 2002, Walter Kim (2002) commented on the ever-increasing number of mental disorders that were recognized by the American Psychiatric Association (APA). "Given so many maladies to choose from," he said, "a person who can't find at least one of his problems covered in DSM-TV must have something really wrong with him" (p. 92).

The Diagnostic and Statistical Manual of Mental Disorders (DSM), now in its fourth edition (DSM-IV) is a mental health clinician's primary tool for identifying and categorizing mental disorder in the United States. It contains a list of all mental illnesses currently recognized by the APA along with detailed descriptions of each disorder and a set of criteria that aid in identification and diagnosis. DSM's influence, however, extends beyond diagnostic boundaries: it guides research into psychoactive pharmaceuticals, it is used by insurance companies to authorize reimbursement for mental health services, and its definitions and disorders often play a role in legal decisions of sanity and culpability. Fundamentally, DSM delineates in our society which behaviors are reasonable and expected in the course of a normal lifestyle and which behaviors constitute the presence of mental illness.

As mentioned, DSM is currently in its fourth edition, indicating that it has not remained the same throughout its existence. In fact, it has at times undergone massive reorganizations and if the first version of DSM were placed next to the most recent edition the two books would scarcely seem related. Revisions and reprinting are not uncommon events in and of themselves; many documents, from textbooks to tax codes to college guides, go through multiple editions over the course of time. There is no reason to expect that classification schemes for mental illness would be any different, for as Grob (1991) noted, "Nosologies - psychiatric or otherwise - are rarely etched permanently in stone" (p. 430).

Bearing this impermanence in mind, the purpose of this paper will be to explore the reasons behind change in psychiatric nosology, and particularly DSM, throughout the history of American mental health care. If DSM is simply a compendium of science based knowledge regarding mental illness, accruing more knowledge and drawing closer to a definitive system with each edition, then this would simply be a history of the accumulation of formal psychiatric knowledge. As it would happen, this is not the case. As Bayer and Spitzer (1985) explained:

For many years, philosophers concerned with the explication of the logic of scientific inquiry portrayed an image of scientific progress that disregarded the extent to which social and professional interests came into play. Scientific activity and controversy are now understood to be affected in important ways by intraprofessional interests, as well as by broad historical and social trends (p. 195).
Consideration of these professional interests and socio-historical trends place the DSM into a much broader focus, illustrating the various events and transactions - both inside and outside of psychiatry - that have motivated revisions of psychiatric nosology and revealing some of the implications these alterations had, and continue to have, on the process of understanding and diagnosing mental illness.

The Creation of DSM

Grinker and Spiegel’s sentiments, written in 1944, were shared by a great many of WW II’s returning psychiatric professionals and would ultimately shape the construction of a revised clinical nosology. As previously mentioned, these returning individuals constituted nearly half of the total number of psychiatrists in America, and they were a group eager to apply the psychiatric lessons they had learned in war to mental health practices at home. This “group” of like-minded practitioners, however, was at the moment informal; they were not unified under any common heading. Without some sort of organization, it is unlikely that such a diverse population could have enacted a lasting change in the profession of psychiatry.

An organization did emerge, though, under the leadership and influence of William Menninger. In 1946, the Group for the Advancement of Psychiatry (GAP) was founded at a meeting of the APA, the organization that the GAP was determined to reshape. Grob (1991) summarized the thrust of the movement well in saying, “The new organization [GAP] was based on the presumption that psychiatry’s responsibilities and functions transcended institutional care and treatment of the mentally ill. ‘I do feel,’ Menninger told a colleague in early 1947, ‘that American psychiatry needs renovation in the sense of consideration of social problems and social needs’” (p. 428).

Though it cannot be said with certainty, one can conceive that Menninger called it a “renovation” and not a “revolution” out of hesitance to invoke such language in the wake of a World War, rather than as the most adequate portrayal of the circumstances surrounding the creation of the GAP. Menninger envisioned a sweeping change in the aim of professional psychiatry, much more than a simple retooling of what was already present. From the rhetoric of visionary mental health professionals of this era, a sense of far-reaching social activism emerges, and that is exactly what Menninger and the GAP were advocating.

Little time was wasted in beginning the nosological revision: in 1948 the APA Committee on Nomenclature and Statistics opted to postpone changes in its current manual (i.e. *Statistical Manual for the Use of Institutions for the Insane*) and instead sought suggestions for a major change. By 1950 the Committee had prepared a draft of the new psychiatric nosology, an amalgam of three nomenclatures in use by the close of WWII: the APA’s Standard Classified Nomenclature of Disease (the *Statistical Manual*), the Armed Forces Nomenclature (War Department Technical Bulletin, Medical 203), and the Veterans Administration Nomenclature (only slightly different from Medical 203) (Hours, 2000, p. 945).

The process of revision took only two years. The 1950 draft of the new manual, titled the *Diagnostic and Statistical Manual: Mental Disease (DSM)*, was circulated to various individuals and organizations within the psychiatric profession and also to a sample of 10% of the APA membership (selected to represent a cross-section of geographic locations, occupational settings, and organizational memberships). Of that ten percent, 46% returned the 9-page questionnaire accompanying the draft, and 93% of these respondents were said to have approved the initial draft (Houts, 2000, p. 945). A second draft was prepared utilizing this first set of revisions, and this draft was approved by the APA membership and its governing body in 1951; the manual was published and distributed by the APA the following year.

In *DSM*, mental illness was divided into three broad, logical categories: organic brain syndromes, functional disorders, and mental deficiency (retardation). Comparatively, the *Statistical Manual* had twenty-two, rather specific categories. Additionally, *DSM* was much less dominated by psychoses and contained greatly expanded sections on psychoneurotic reactions, personality disorders, and transient situational personality disorders. Finally, as the name of the new manual suggested, it was not only oriented toward the facilitation of gathering statistical information but also towards professional diagnosis. George Raines (1953), chair of the committee responsible for *DSM*, explained why:

"It has become popular to decry diagnosis of psychiatric illness with the contention that each patient is an individual so different that standard labels cannot be applied. Without sound diagnosis, statistics are inaccurate and misleading, or unavailable; factual data cannot be accumulated from past experience to guide the future; and accumulated knowledge is transmitted with great difficulty, if at all (Raines, 1953, p. 548).

Ultimately, and in contrast to the governmental influences behind the previous standard nosology, *DSM* is best seen as a document influenced by sweeping changes in both the professional and social ideology of psychiatric practitioners in the years after WWII. Professionally, practitioners switched from a primarily medical, biological orientation to a chiefly psychoanalytic view as a result of their experiences in WWII; this change is mirrored in the greatly increased prominence of psychodynamics in *DSM*. In terms of social ideology, psychiatrists became much more interested in applying mental health practices to non-institutional populations as a result of growing social activism and prevailing post-war ideas of social betterment. William Menninger (1947/1967) explained this social activism by saying:
Psychiatry should place high priority on its efforts to provide the "average" person with psychiatric information he can apply to his own problems. The public wants this education. Very possibly it may increase the number of patients who seek help from a psychiatrist, just as a campaign about cancer or tuberculosis increases the number of patients who go to doctors about these problems (p. 579).

Menninger's last sentence, in particular, highlights one of two caveats to the seemingly altruistic social activism proposed by post-war psychiatric professionals. This "possible increase in the number of patients" seeking psychiatric services was not only an off-handed coincidence of an increase in education, but also a professional necessity if the huge number of psychiatrists returning from the war were to have any hope of working. A doubling of active practitioners in psychiatry, which WWII generated, would need (at least) a doubling of patients in order to sustain that many professionals.

Secondly, the years after WWU saw a huge rise in the amount of federal funding available for training and research in psychiatry and clinical psychology. For example, as a result of the National Mental Health Act of 1946, the amount of fund dispersed by the NTMH between 1948 and 1962 for research and training grants rose from $374,000 to $42.6 million and from $1.1 million to $38.6 million, respectively (Starr, 1982, p. 346). Clearly there was a lot of money available within various areas of mental health care, creating strong incentives to remain in a profession that previously had been in the doldrums of the medical world.

**DSM-II: The First Revision**

In contrast to DSM, the first official revision of this manual, DSM-II, was not accompanied by the broad social and professional changes that shaped DSM. As Kutches and Kirk (1997) explain, "making DSM-II was a relatively private and simple process, more like changing rules and regulations within one organization than negotiating treaties among many rivals factions, each with a very different objective" (p. 40). The rather modest goal of the work, according to two of its consultants, Spitzer and Wilson (1968), "is the improvement in treatment and prevention that will evolve with better communication among psychiatrists of all nations" (p. 1929).

Assisting international communication, in fact, was the main motivation for changing DSM. The *International Classification of Diseases (ICD)* had long been published by the World Health Organization (WHO) to facilitate the collection of disease statistics, including those on mental illness around the world. In 1965, asked its Committee on Nomenclature and Statistics to revise DSM in order to make it compatible with ICD-8. The new diagnostic manual would also include changes in professional concepts of mental disorder that had occurred since 1952, namely the inclusion of new disorders and the exclusion of unused ones. Gerald Raines, chair of the committee that produced DSM, had anticipated such revisions:

Sound diagnosis is possible only with a nomenclature in keeping with current concepts of psychiatric illness, and sufficiently flexible and inclusive to permit the introduction of new ideas. It is therefore necessary that psychiatric nomenclature, as well as other medical nomenclature, be revised from time to time to keep it in step with the moving body of psychiatric knowledge (Raines, 1953, p. 548).

The actual process of revising and changing DSM was quite simple, and relatively isolated from outside influence. Whereas DSM organized mental disorders into three categories, DSM-II contained ten such categories. The categorical increase was chiefly the result of the reclassification of six of the mental illness clusters collected under a single DSM heading ("functional disorders") as individual categories to indicate how mental disorder was often viewed in America. DSM-II also included a new section on childhood disorders ("behavior disorders of childhood and adolescence") that did not appear in DSM. All told, fifty-four diagnoses were added to DSM-II, raising the total from 108 in DSM to 162 in DSM-II, beginning a trend of increasing the number of diagnoses that has continued through the most recent edition (Houts, 2002, p. 20).

Perhaps the most notable occurrence in DSM-II, in that it foreshadowed greater changes to come in later revisions, was the elimination of the term "reaction" from the text. "Schizophrenic reaction," for example, became simply "schizophrenia." According to the Committee on Nomenclature and Statistics, this was done in order:

... to avoid terms which carry with them implications regarding either the nature of a disorder or its causes... In the case of diagnostic categories about which there is current controversy concerning the disorder's nature or cause, the Committee has attempted to select terms which it thought would least bind the judgment of the user (APA, 1968, emphasis in original).

This move towards an atheoretical perspective was for the time being incomplete, as several diagnoses still retained the label of "reaction" ("psychotic depressive reaction," for example). Curiously, the category on childhood and adolescent disorders - the only entirely new set of diagnoses in DSM-II - was composed entirely of "reactions."

Taken as a whole, DSM-II was only subtly different from DSM. The major factors that had influenced the creation of DSM (i.e. psychodynamics and shift in focus away from practice in mental hospitals) remained intact in its first major revision. There were hints in its production, however, of greater changes to come. Spitzer and Wilson (1969), replying to various criticisms of DSM-II, suggested that future revisions involve various subcommittees to supplement the actions of the small committee.
charged with producing the next edition. In the same article, Spitzer and Wilson again noted, perhaps most presciently of all, that, "McCall and Jackson [two who critiqued the volume] make strong cases for considering the neuroses to be essentially symptom complexes devoid of any independent existence as nosological entities. Although we believe these ideas have considerable merit, they will have to convince the profession of their approach before they can be incorporated into an official classification system" (p. 424).

**DSM-IV-TR and Beyond**

DSM-IV was released in May 1994, amid much coverage in the national media. Receiving far less attention nationally was the release two years later of a DSM-IV Coding Update that included numerous modifications in the code numbers assigned to particular mental disorders. Within this slim coding volume was an announcement of intentions to release a text revision of DSM-IV in 1999 or 2000, to include no alterations in diagnoses or criteria sets, but instead to make simple additions to sections providing information on course, prevalence, etc. Kutchins and Kirk (1997) recalled with suspicion that, "Very shortly after publication of DSM-III, a revision was announced and users were assured that it was to be a minor midcourse correction to incorporate new findings and rectify errors. When DSM-III-R appeared, most of the diagnoses had been changed, new ones had been added, and other substantial alterations had been adopted" (p. 265).

The revision of DSM-IV (DSM-IV-TR) was released in 2000 to little fanfare. In contrast to the liberties taken by the Work Group to Revise DSM-III, the group in charge of DSM-IV-TR stayed within the boundaries of a minor text alteration, adding only 50 pages to the manual and no new diagnoses. Its goal was simply to facilitate early inclusion of growth in research knowledge that had occurred, because DSM-V was not scheduled for print until at least 2010.

This timeline for DSM-V is still intact, with no hints of any changes in the schedule of release. The greater psychiatric profession, however, is already preparing for the next revising process, and a current search for "DSM-V" in PsycINFO yields over 35 hits for articles containing both suggestions for the manual and predictions of DSM-V’s outcome. Barring any major paradigm shifts in the psychiatric profession, DSM-V will likely build on the evidence-based tradition that Frances firmly established in the creation of DSM-IV. In fact, it is likely that until research psychiatrists lose the hold on the profession that they gained in the 1970’s with the fall of the psychodynamic orientation, future revisions will continue to parallel the current research-oriented trends of the mental health care world.

Ultimately, as psychiatry begins to uncover more and more of the “biological abnormalities” to which Kendell (1991) referred, a major shift in psychiatric nosology is liable to occur. This move will be to a diagnostic system that moves away from the current emphasis on symptom-based diagnostic criteria and towards etiological descriptions of mental disorders (Schaffner, 2002). As Kilhström (2002) entreats:

For Kraepelin, diagnosis by symptoms was a temporary fallback, to be used only because diagnosis by pathology and etiology was not possible. This “fallback” has dominated our thinking for more than a century, and it is time to press forward, with all deliberate speed (p. 297).

Such a perspective is not only advocated by Kihlström (2002) and Schaffner (2002) but also, more subtly, by those producing each DSM edition, who hail each subsequent version as more efficacious and more valuable than the last. Grob (1991) addressed the idea of progress in nosology in the conclusion of his essay on the history of DSM:

To contemporary psychiatrists the history of nosology represents aversion of the idea of progress; advances in knowledge supposedly lay the foundation for the creation of new categories that describe reality in better and more accurate terms. Within such a perspective, the history of psychiatry is moving on an upward gradient toward an ideal end. In this specific instance, the final goal is a definitive and presumably unchanging nosology of mental illnesses (p. 430).

It is difficult to argue with the desire for such an outcome, where all issues of etiology, diagnosis, and treatment can be comprehensively addressed with the publication of a single volume, however unwieldy. However, the want for such an outcome precludes an understanding of the history and development of psychiatry and psychiatric nosology; insights provided by this professional meta-narrative illuminate the danger in considering the eventuality of an ideal, unchanging psychiatric nosology.

There is the fallacy, often implicit in the claims of any profession (not just psychiatry), assuming that all progress is progress towards something better - in the case of psychiatric classification, an ideal, static nosology. In the case of psychiatry, this is not truly a fallacy, because it is relatively unarguable that the profession’s nosology has progressed since Greek humeral theory and even since the first official nosology, DSM. However, one must take care not to equate progress within the limitations of the surrounding professional environment with advancement towards an ideal end. This is a mistake often made in the appraisal of biological evolution, where natural selection can easily be misconstrued (with only a small touch of anthropocentrism) as the means directed at achieving an “optimal” end.

Beutler and Malik (2002) applied this idea to DSM’s using the evolutionary arguments of R.A. Fisher and Sewall Wright (see Provine, 1986), who argued back and forth regarding
whether or not the majority of current species had reached their full evolutionary potential. Fisher asserted that species were mostly optimized, residing upon what could be imagined as the highest peak, surrounded by valleys of lower evolutionary fitness. Wright countered this by proposing that the peak on which most species resided is not likely the highest peak in all the landscape but rather what he termed a "local optimum." This would represent the highest point that a species could see from where they were now, though not necessarily the highest point possible. Psychiatric nosology, Beutler and Malik (2002) asserted, is likely subject to the same shaping forces and the same tendency to reside on a local optimum.

As this thesis has examined, the profession of mental health and its practitioners have been prone to fluctuations throughout their history, exposed to numerous shifts in the loci of power, social and technological experience, theoretical orientation, and financial resources - the shaping forces that Beutler and Malik (2002) referred to. These professional alterations affected not just the day-to-day activities of mental health practitioners but also the predominant psychiatric nosology, resulting in the substantial changes in classification that have been discussed throughout this paper.

If this history of change affords us any glimpse of the forthcoming, then the profession of psychiatry is unlikely to exist as a static entity future. Additionally, based on historic trends, it is just as unlikely that future shifts in the psychiatric profession will be unaccompanied by alterations in the nosology of mental illness. The idea of progression towards a definitive, unchanging nosology seems a tenuous outlook to endorse.

Much less unreasonable is the belief, outlined earlier, that nosology will eventually move towards a focus less on symptomology and more on etiology and ultimate causation. As research into mental disorders progresses and knowledge of genetic and environmental contributions to psychopathology accumulates, it is hypothetically possible that later DSM editions could ultimately become objective, irrefutable sources of knowledge on classification and diagnosis, a sort of Periodic Table of mental illness. The notion of a DSM Task Force setting out to revolutionize the classification of mental illness could be laid to rest, for the only changes possible in a definitive nosology seems a tenuous outlook to endorse.

The ultimate feasibility of this outcome for psychiatric nosology is highly and, for the moment, irresolutely debatable. What should be emphasized, however, is that psychiatric nosology does not emerge simply as the result of "advances of knowledge" (Grob, 1991, p. 430). Rather, systems of classification are born of a myriad of issues and influences that circumscribe the knowledge itself: whose knowledge is paid attention to (professional prominence), changes in where the knowledge comes from (locus of professional power), who has the funds to produce the knowledge (financial resources), how the knowledge is interpreted (theoretical conceptions), and the environment of knowledge application (social/technological surroundings). In this light, even the presence of a complete, objective knowledge base regarding mental disorders may not guarantee a definitive, unchanging nosology. As Grob (1991) noted, perhaps, "the only constant is the process of change itself. The search for a definitive nosology, therefore, may simply be an expression of the perennial human yearning for omniscience - an attribute eagerly sought by many but never yet found" (p. 430).

End Notes:

1 Bearing in mind that there are two prominent APA's in the realm of behavioral science (American Psychiatric Association and American Psychological Association), it will be used in this paper only to refer to the American Psychiatric Association.

2 "Nosology," a term frequently used in this paper, is a classification or list of diseases

3 The ten categories in DSM-IV are: mental retardation (all causes), organic brain syndromes, psychoses not attributed to physical conditions listed previously, neuroses, personality disorders and certain other non-psychotic mental disorders, psychophysiological disorders, special symptoms, transient situational disturbances, behavior disorders of childhood and adolescence, and conditions without manifest psychiatric disorder.

Works Cited:


Published by ScholarWorks@UARK, 2003
Faculty mentor comment:

Dr. David Schroeder, in his letter to the Inquiry Publication Board, made the following remarks about Mr. Jackson’s work:

Eric Jackson’s paper, “Organizing Madness: Psychiatric Nosology in Historical Perspective,” is based on Eric’s Honor’s thesis and provides a unique review of the multiple influences that have shaped psychiatric classification schemes, including the most recent edition of the Diagnostic and Statistical Manual. Mental Disorders (4th Edition, 1994). Eric traces the evolution of diagnostic manuals from the earliest attempts of psychiatry and psychology to model medicine’s diagnostic systems for physical diseases for the creation of a similar nosological scheme for mental diseases. While benefits for patients were undoubtedly realized by these classificatory advances, Eric explores other factors that are largely unrelated to treatment per se (e.g., professional authority, financial consideration) as important roles played in the development of diagnostic manuals. While these factors certainly continued to impact revisions of early manuals, and ultimately the DSM, and the continuing series of revisions, Eric also integrates the influences of changes in the dominant psychiatric and psychological orientations (e.g., the shift from a Freudian psychodynamic view that held sway in the early part of the 20th century to more objective and behavioral-based views of abnormal behavior that emerged with the “cognitive revolution” that began in the 1960’s and continues today). These changes not only led to changes in terminology (e.g., the elimination of the term “neurosis” from the psychiatric lexicon) but also to the explicit specification of defining an empirically validated symptom pattern for each diagnostic category.

Eric has offered a masterful integration that brings together works form psychiatry, psychology, medical anthropology, and the history of medicine to provide the context for his review. He provides a glimpse “behind the scenes” at how the most important reference for psychiatry, clinical psychology, and other mental health professions has been shaped. His submission is extremely well written, and he presents his analysis without bias or preconception. This paper is a strong piece of scholarship that does not lose the reader in professional jargon. It should be an interesting and enjoyable “read” for people from a wide range of backgrounds.
HISTORY:

Karilyn Moeller. Chinese Women Unbound

CHINESE WOMEN UNBOUND:
AN ANALYSIS OF WOMEN'S EMANCIPATION IN CHINA

by Karilyn Moeller
Department of History

Faculty Mentor: Dr. Henry Tsai
Department of History

Abstract:

"Chinese Women Unbound" gives a brief historical background of the status of women in China and presents a well-documented history of the evolutionary process of Chinese women's emancipation—from the first missionary school for girls in the 1840s, to the first females admitted to Beijing University in the late 1920s, the marriage law of 1950, and the divorce rate in the 1990s, among other events. The paper also discusses Chinese women's involvement in the 1911-1912 revolution, the Communist revolution, and the modernization of Chinese economy. In narrating this evolutionary process, Moeller analyses the various forces behind the changes, as well as the social, cultural, and political issues that were intertwined with the women's movement in China. The original version is 24-pages long; the article presented here is a condensation made for this publication by the author.

"For all women there are the three obediences—to the father before marriage, to the husband after marriage, and to the son after death of husband..."

Yu-fang was born in 1909 in southwest Manchuria. She was the first daughter of her fifteen-year-old father and twenty-one-year old nameless mother. At age two, Yu-fang’s feet were bound in the popular fashion by bending the toes under the sole, wrapping tightly with cloth and crushing the arch with a large stone. Yu-fang’s father had one valuable asset, his beautiful daughter Yu-fang. From the time of her birth he had plans to use this asset to better himself. Yu-fang was groomed to be a high-class courtesan. When she was fifteen, her father sold her as a concubine to General Xue Zhi-heng who was connected to the warlord government in Peking. Yu-fang did not agree with the arrangement, but to resist her father would have made her a very unfilial daughter. Nine years later Yu-fang faced another terror. General Xue was gravely ill, and she told her granddaughter years later, “if the general died she would be at the mercy of the wife, who had the power of life and death over her ... sell her to a rich man, or even into a brothel, which was quite common”.

Yu-fang’s early life was very typical of the oppression and subjection Chinese women endured for thousands of years. The depravity of women taught by Confucian ethics was the basis for this subjection. This paper will briefly discuss the status and condition of women in dynastic China in the areas of marriage, property, education and the practice of footbinding. Events and circumstances that brought about changes in the lives of women and when those changes began to take place will also be examined, as well as an examination of the role and status of women in China today, especially in the areas of marriage, education, employment and participation in government.

Confucian philosophy and ethics dominated Chinese family and cultural dynamics after Han Wu Di (156-87 B.C.), Emperor of Western Han Dynasty, adopted a policy to “Reject all other schools of thought and hold only Confucianism in esteem.” The patriarchal and patrilineal family structure dictated by Confucian ethics taught the inferiority of women and became the basis for sexual discrimination and subjection of Chinese women and girls. According to Confucius, a virtuous woman had three obediences: to her father and brothers before marriage, to her husband after marriage, and to her son after death of husband. Early Chinese marriage practices and customs exhibited many elements of a male dominated patriarchal family structure. The “bride price” paid by a groom’s family was indicative of the idea that females were “property” to be bought and sold. After a young bride moved into her husband’s family home, she was subject not only to him, but to other members of her new family. A woman’s duty to her parent’s-in-law was the subject of teaching in the Book of Filial Piety for Women. The subject of serving parents-in-law stated:

With regard to a woman’s service to her parents-in-law, she is as reverent as to her own father, as loving as to her own mother. Maintaining this attitude is a matter of duty, and adhering to it is a matter of ritual. When the cock first crows, she washes her hands,
rinses her mouth, and gets dressed to make her morning call. In the winter she checks that [her parents-in-law] are warm enough, in the summer cool enough. In the evening she checks that they are settled, in the morning that they are getting up...  

If a wife failed to bear a son, a man could bring a secondary wife or a concubine into the household. A concubine was usually acquired by purchasing a young girl from a poor family. According to Florence Ayscough in Chinese Women: Yesterday and Today, "ancestors required male descendants. If one woman did not produce them, another should be given the opportunity to do so." A concubine's duty was to serve the first wife and bear children. Whether taking a concubine for reproduction or as an object of pleasure, the practice was degrading and humiliating for women.  

In the Chinese kinship system, divorce was nearly impossible for a woman. A man could divorce his wife for any one of seven reasons, but women had no rights in regard to divorce, separation or property. The fate of concubines was even more precarious. They could be expelled by the husband at any time and also by his family after his death. 

A widow was obligated to continue serving her dead husband's family. Confucian norms were that she would remain a widow, but her in-laws had the right to sell her if they chose. Any children she bore into the husband's family legally belonged to the family. The practice of immolation was quite common and widows were exalted if they "followed their husbands on death." Immolation for many widows was more desirable than the prospects of continuing with the dead husband's family. 

One of the most graphic examples of the restrictions Chinese women had to bear was foot binding. Began during the Song Dynasty (960-1279), the practice of binding young girls' feet so they would be tiny little "golden lilies" was touted as necessary in order to obtain a good marriage. But, many young girls did not survive the effects of rotting flesh and broken bones. The physical act of binding feet was confining, and women with bound feet experienced physical pain and limited mobility throughout their lifetimes. In a Confucian culture, foot binding was a symbol of dependence and subservience. 

The movement against foot binding began during the Manchu rule in 1644. According to Alison Drucker, after China's ports were opened to western missionaries by the 1842 Treaty of Nanjing, pressure increased to stop the practice of foot binding. Chinese converts were encouraged to stop the practice, and western opinion that the practice was barbaric also had influence. Mission boarding schools from 1867 on refused to keep girls unless their feet were unbound. The T'ien tzu hui (Natural Foot Society) was created in 1895 by ten foreign women, but by 1908 it was operated exclusively by Chinese, including a male Chinese doctor. This society was very active in promoting natural feet, even getting an audience before the Qing court. The missionary societies had some indirect influence on abolishing foot binding, but only after the efforts became more Chinese and nationalistic was there any success. 

Leaders of the Taiping Rebellion (1850-1864) not only prohibited foot binding on pain of death, but also opened the door of emancipation for women in several other areas. Intellectual Kang You Wei, leader of the 1898 Hundred Days' Reform, was an outspoken critic of foot binding. By 1905 girls attending state schools could not bind their feet and by 1908 many Chinese intellectual and political leaders were speaking out against the practice. Many of these leaders had been exposed to western education. Yuan Shi Kai, a social conservative, publicly announced foot binding as injurious to women's health and an impediment to their education. Chang Chih-Tung, governor-general of Hupei and Hunan from 1889-1907, also reiterated Yuan's sentiments. Foot binding diminished the productivity of females, both inside and outside the home. Foot bound women could not enrich their families or their nation. The Nationalist government that came into power in 1911, banned foot binding outright. A survey done in 1929 in an area approximately 125 miles south of Peking revealed that 99.2 percent of the women born prior to 1890 had bound feet. None of the women born after 1919 had bound feet. 

On the education of women Confucian thought was "...women indeed are human beings, but they are of a lower state than men, and can never attain to full equality with them. The aim of female education therefore is submission, not cultivation and development of the mind." The thrust of women's education was how for women to properly fulfill their role as virtuous wives and good mothers. There is however, evidence that many girls of the elite class were taught to read and write. The Book of Filial Piety for Women written around 730 by Nee Zheng and translated in Under Confucian Eyes is an instruction book on how to be a good filial woman. That Miss Zheng wrote such a book is evidence that some women could read and write. Under Confucian Eyes also has examples of other Chinese women writers from the Ming and Qing eras. Ban Zhao (A.D. 41-ca.115) is another example of the flexibility of the Confucian system. China is indebted to Ban Zhao for the completion of the history of the Han dynasty left unfinished after her brother Ban Gu, a well-known historian, died. She was also author of the text Precepts for Women. 

Educated women in early China were the exception. Few women possessed more than basic literacy, and peasant women would not have had even that. Early missionaries reported they seldom met a Chinese woman who could read. One female missionary reported, "With very rare exceptions women are never educated. Of heathen women possibly one in two or three thousand can read." From Chinese, comments like, "can you teach the horse to read and write" and "oh, but the women . . . they can't learn" were common. Girls usually married in their
teens and became the property of another family, so the expense of educating girls was thought a waste of resources.21

The first school for girls in China was established in 1844 by Mary Ann Aldersay in Ningbo. Within a year Miss Aldersay had fifteen students and by 1852 she had forty. Within the next fifteen years schools opened in Shanghai, Foochow, Canton and Amoy, all under the direction of various mission organizations. The Treaty of Tianjin in 1858 opened all of China to missionaries. Thereafter the opening of girl’s schools began to move inland and north. Schools were opened in Tianjin and Peking in 1864 and Chefoo in 1872. By the early 1900’s mission schools had more potential students than they could accommodate. These schools provided education free of charge for girls from the poorer sector of society. Many of the students from these early mission schools became teachers in the governmental institutions established early in the 20th century.22

Late in the 19th century, Chinese intellectuals began to be more vocal about women’s issues, especially the education of women and girls. Liang Qichao in 1897 expressed his views in an article:

In China today, whenever the subject of women’s education is debated, someone is bound to say, “There are much more important and urgent issues that this business of women’s education”... However, I think the cause of weakness and failings in our society can be traced to the fact that women’s education in this country has long been ignored.23

Growing sentiment such as this caused the Chinese to take some initiative by establishing private schools for girls. The first was established in Shanghai in 1898. This school was the forerunner of many private schools that were established in the first decade of the 20th century.24 By 1907 in Shanghai alone, there were over 800 girls attending schools established and funded by private Chinese citizens.25

Empress Dowager Cixi issued an edict in 1901 permitting the establishment of government-sponsored schools for girls. Nothing was officially provided until 1907. Boys and girls attended school together in lower elementary grades, but had to have separate classes in higher primary schools.26 After the founding of the Republic of China in 1911, females in primary education made up only 2 percent of the primary school population. This figure increased to 19.2 percent by 1936. The percentages for secondary education rose from 9.8 percent in 1911 to 17.6 percent by 1930. Figures gathered in 1916 give a better perspective on how young female education was. Estimates are that 95 percent of school age girls were not in school.27

Opportunities for higher education for women before the May Fourth Movement (1919) were available only through three all-woman universities run by foreign churches. Female secondary school students participating in the May Fourth Movement became torch-bearers for women’s liberation and equality in China. They had a voice that could no longer be ignored. Cai Yuanpei, chancellor of Beijing University, and Dr. Sun Yat-Sen, the founding father of the Republic of China, were both driving forces behind the advancement of women’s education. On May 6, 1921, Sun, speaking at a girl’s school in Guangdong Province, remarked, “there would be no equality between the sexes without women being educated, and... without sexual equality there would be no representative movement.”28 Cai, as chancellor of Beijing University, admitted two female students in 1920. Those students, however, were not been allowed to take the entrance examination and were considered “visiting” students. Cai was publicly defiant, declaring that the university charter did not prohibit the admission of females and there was no reason to refuse them. Cai stated “the road to our country’s prosperity is through universal education: in order to achieve this, one should begin in earnest with women.”29 Females were officially enrolled in Beijing University in the late 1920’s. After that females were admitted into secondary schools and universities across China. This was the beginning of higher co-education in China.30

Concurrent with education, was progress in other social issues. Individuals who study Chinese history often credit the Taiping Rebellion of 1850-1864 as the beginning of emancipation for Chinese women. Hong Xiuquan, the Hakka leader of the rebellion wanted to establish a new dynasty where women would be equal with men. Under the Taiping government women served in the military as generals and soldiers. Women served with courage and in large numbers. A Qing official wrote about an upcoming battle over the control of the city of Nanjing; “after we recapture the city, all the Guangxi women should be executed. Absolutely no leniency or mercy should be shown them. For they have been just as courageous and fierce as male soldiers in defending the city.”31 The Taiping government also outlawed foot binding, concubinage and prostitution. The traditional practice of arranged marriages was replaced with free choice monogamous love matches. In addition, land was distributed to women and men equally. Although Qing forces put down the rebellion, a new era for Chinese women had been ushered in.32

The leaders of the Self-Strengthening movement, which followed on the heels of the failed Taiping Heavenly Kingdom, began to see women in light of industrialization and modernization. Women were being recognized as potential factory workers. Zhang Zhidong, governor of Hunan and Hubei provinces in the late 19th century, used the terms “essence” and “practical use” to develop China’s first industrial area. Zhang’s goal was to retain the “essence” of Chinese moral and philosophical values, (in this case, women’s place was in the home) yet put to “practical use” western knowledge and technology.33 Nonetheless, Zhang begin to see women in light of what they could contribute to Chinese industrialization. By the late 19th century a large number of women and girls were employed in the silk-reeling industries around Wuhan and
Shanghai. Their small agile fingers were especially suited for this work, and it quickly became difficult to hire men to replace women because it was considered “women’s work.” Naturally, women responded to their new role by becoming more independent and assertive.

In contrast to the Self-Strengtheners “essence” of keeping the old Chinese ways, proponents of more drastic change came forward in an 1898 movement known as the Hundred Days Reform. One of the leaders, Liang Qichao, was an advocate for women’s liberation and education. Liang, a well known journalist, wrote that China’s poverty was due to the fact that women were totally dependent upon men. He described the dependent, illiterate women as parasites of Chinese society, consuming but not producing.

Women’s reform movements in China often paralleled national reform movements. Such was the case with the Red Lanterns, the female counterpart to the Boxers. The Red Lanterns was an organization of young girls between the ages of twelve and eighteen. Their participation in the burning and killing that took place during the Boxer Rebellion of 1898-1900 reveals that young women had passion and were no longer going to be spectators in the reformation of China. Those young girls had the courage to become participants. Many lost their lives but they made it possible for future generations of Chinese women to gather the courage to fight for their rights. A piece of propaganda from the Red Lanterns read: “The red lantern shines, lighting the path for the people.”

From 1900 to the present, Chinese women have been involved in revolution. Early in the 20th century, women’s magazines began publishing revolutionary propaganda for women. Chen Xiefen, a revolutionary writer, argued that men’s promotion of women’s rights and education had been strictly for the benefit of man and the state. It was time for women to promote themselves. Qiu Jin (1875-1907) has been labeled Woman Revolutionary, and is a hero to women in China. Qiu Jin had bound feet and was forced by her parents into an unhappy marriage. She left her husband and children and went to Japan to study. After returning to China, she became active in secret societies whose aim was to overthrow the Qing dynasty. She also spoke to women’s groups challenging them to take control of their future. In one speech she told the women:

... If you have children, please send them to school by all means. Girls, no matter what, never have your feet bound. Young women, if possible it’s best for you to go to school; but even if you can’t, then read at home and study your characters all the time ... Everyone, the nation is on the verge of collapse. Men can no longer protect it, so how can we depend on them? If we fail to rouse ourselves, it will be too late after the nation parishes.

Qiu Jin was executed in 1907 at the age of 32, after plans for an armed revolt were exposed.

Early in the 20th century Chinese women were also forming women’s rights organizations. Of the groups studied, all were founded by Chinese women who had studied abroad or mission schools in mainland China. These groups openly opposed concubinage, female slavery, and foot binding, and promoted education and encouraged economic independence. The Free Marriage Lecture Society formed in 1909 in Hunan met with resistance when speaking out against arranged marriages. The society was banned and all governors were ordered to be on the look out for “unseemly” activities.

In 1911, women’s revolutionary activity again paralleled national activity. Women were involved in the revolution that brought down the Qing dynasty. Female military units were formed and women spoke in the streets exhorting crowds to support the revolution. Women in the military also served as barbers: cutting the queues of Chinese men rejecting Manchu rule. Chang Chu-chun, a trained physician, was opposed to women in combat, but she organized forty women to go with her to the front line and tend to the medical needs of the wounded.

Women’s involvement in the 1911 revolution was focused more on nationalism that on specific women’s suffrage. This changed by the mid 1910’s when young educated reformers begin taking up the pen to advocate women’s liberation. Democracy had been achieved, so the focus again turned to social issues. Essays on chastity and arranged marriages began appearing in New Youth, a revolutionary magazine put out by Beijing University faculty. Stage performances of A Doll’s House, where lines such as “don’t become a man’s play-thing” and “demand freedom” became inspirational slogans for young people. When news of the Treaty of Versailles reached Beijing, the youth all over China reacted. Female students as young as thirteen were involved in mass demonstrations. They were demonstrating for the preservation of their country. But their demands continued even after China refused to sign the treaty. Women’s voices grew louder in their demands for equality and liberation.

In China the years from 1919 until 1949 were years of nearly constant unrest. The nation had to deal with the civil wars between the Nationalists and the Communists, as well as the imperialistic aggression of Japan. Women’s social issues were again secondary to the issue of preserving the nation. The establishment of the People Republic of China in 1949 was the catalyst that brought women’s issues back into the mainstream of social reform.

“Men and Women are equal. Women can hold half the sky”

In the year 1919, Mao Zedong wrote a series of articles titled A Critique of Miss Zhao’s Suicide for a newspaper in Changsha. The articles were a response to the recent suicide of a young bride named Zhao Wuzhen. Miss Zhao, in protest of the
The People’s Republic of China (PRC) was established on October 1, 1949 by the Chinese Communist Party with Mao Zedong as the Chairman. The PRC quickly began a mass movement in social reforms, particularly in the area of women’s rights and equality. The first session of the Chinese People’s Political Consultative Conference met in Beijing late in 1949. Ten percent of the delegates present were women. Those sixty-nine women represented Chinese women throughout the country. Women had leading roles in the organization and implementation of the new government. The vice-chairperson of the new Central People’s government was Soong Qing Ling. Since that time, the number of women participating in political activities has continued to increase. The most basic way women exercise their freedom and political rights is to participate in elections. In the first general ballot in 1953, 90 percent of the women in China cast their vote. The number has risen to 95 percent today. After the Fourth World Conference on Women was held in Beijing in 1995, the Chinese government began an aggressive campaign to increase women’s involvement in government and political affairs. Today there are 650 female deputies (21.8 percent) to the Ninth National People’s Congress (NPC), and there are two female vice-presidents of the Standing Committee of the NPC. The State Council has two female ministers and 16 female vice-ministers. In 668 cities across China 463 female mayors or vice-mayors are female. Female cadres account for 35.7 percent of the total across the country. Females account for about 33 percent of all public servants in China.

The participation of women in government from the establishment of the republic to the present has insured the consideration of women’s interests. Very significant is the fact that the first law passed after the establishment of the Republic changed the lives of Chinese women forever. The marriage law issued in 1950 did away with the feudal marriage system of arranged and forced marriages. “Freedom of choice in marriage, monogamy, equal rights for men and women, prohibition of bigamy, concubinage, and child-bride marriages, and freedom of remarriage” became the new marriage law.46 Because the National Marriage Law was such a dramatic change form feudal practices and thinking, a series of educational campaigns were carried out between 1950 and 1953. Doubtless, this made a significant difference in the lives of young women. But, as late as the 1980s there were gaps between principle and practice. Surveys done in 1979 in rural counties in Anhui province revealed that 75 percent of marriages were first negotiated by parents and 10 percent were arranged by parents according to old traditional ways. This was not just a rural phenomenon. In Shanghai during the same year, only four of twenty-two marriages were based on “free choice.”47 In the book Personal Voices: Chinese Women in the 1980s, written by Emily Honig and Gail Hershatter, several young Chinese girls share how their families imposed feudal marriage practices upon them. Today the issue of free choice in marriage, especially in minority and rural areas, is one of the issues being addressed by the All-China Women’s Federation.

Under the Marriage Law divorce also became an option for women. Previously, only men could instigate divorce. This was quickly embraced by women, and if statistics can be used as an indicator, hundreds of thousands of women had been locked into unhappy, undesired marriages. The number of divorce cases brought before the People's Courts in 1950 was 186,167. For the first half of 1953, the number was 398,243. Of the total, 75 percent of the cases were instigated by women. Divorce would have been very difficult for most women, but the land reforms that gave women property made it possible for some to support themselves. There was much resistance from men and mothers-in-law; for losing a wife was like losing a valuable piece of property, and for a wife to willingly leave was like treason. As one saying put it, “a good wife hangs herself, while a wicked woman gets a divorce.” In the early 1950s, thousands of women were killed over divorce and property rights issues.48 Divorce rates had dropped significantly by 1980, to only about 2.66 percent in the city of Beijing. It was much lower in the countryside. In the 80s divorce petitions were sent to a mediation committee. If differences could not be settled then a divorce was granted, regardless of who initiated the petition.49 The national divorce rate in China as of March 2002 is about 10 percent. It continues to be much lower in rural areas. A divorce can be obtained in a day if a couple has agreed on all property division issues. The long term affects this might have on the stability of families and women has yet to be seen.50

Chinese women have also gained the right to own land and inherit property. This was a major advancement in women’s efforts to become self-sufficient. Now, by law, women’s names have to be registered on the family land deed. The land reform in the early days of the Republic distributed land equally to women and men. What this meant to women was expressed by a Shanxi peasant woman who stated, “Our husbands regard us a some sort
of dogs who keep the house . . . that is because for a thousand years it has been . . . after we get our share [of land] we will be masters of our own fate.\textsuperscript{51} Today, women have equal rights with men in ownership of property, allotment of farmland, and inheritance rights. The impact of this change has helped women become more secure socially and economically. Women are no longer confined to the home as consumers. They have become productive, contributing citizens of a growing market economy.

Women workers in China make up 44 percent of the total work force. In rural areas they represent one-half of the work force. The contribution these women make to the economic reform and development of China is immeasurable. In urban areas women are employed in an array of sectors, including education, finance, public health, radio and television, government and social institutions, and social welfare. The growth rate of women employed in these areas from 1982-1990 exceeded that of males by 21 to 78 percentage points. Women with professional and technical titles accounted for 36.8 percent of the total number of persons in professional and technical fields in 1993.\textsuperscript{52}

Rural women have become an indispensable asset to the rural economy. Women generate between 50 and 60 percent of the total rural output value. There are also about 4.8 million rural women self-employed in commerce and service trades. Rural women are the driving force behind the development of township enterprises.\textsuperscript{53} One example is the Wu Guci Village, Tian Yuan Township in the Sichuan province. Chang Yu Xheng, the administrative leader of the village, is a typical peasant woman. Born in 1944, she had seen many changes in her village. But, she reports, men and women are equal and there are many women leaders. She also stated that because women were so involved in work outside the home, men shared in the housework. In addition to farming, the village collectively operates other small industries such as a furniture factory and a shoe factory.\textsuperscript{54} More examples of peasant women’s involvement in township enterprises can be found in Halls of Jade, Walls of Stone by Stacey Peck. These township enterprises in rural China earn the most foreign exchange for the country. Women make up the bulk of the work force in producing textiles, silk, tea, toys, electronics, and embroidery. In Longzhou City, Shangdong Province, the embroidery articles produced by women bring in US $2.5 million annually.\textsuperscript{55}

How has the mass movement of Chinese women out of the home into the work place affected their daily lives, and what obstacles must they face and overcome to experience equality in the work place? Pregnancy and childcare, equal pay, equal opportunity and division of housework and family responsibility are just a few of the things a contemporary working woman would face. It is not possible to address every issue, but a significant issue is childbearing. The Chinese government has been especially diligent in providing protection for employed pregnant women. In urban areas, surveys show that nearly 85.3 percent of child-bearing workers get at least a three-month paid maternity leave. Many state-owned work places provide healthcare, nurseries, and kindergartens. This relieves a great deal of anxiety from the working female.\textsuperscript{56}

Another issue that would be interesting to explore in depth is how Chinese women feel about being such a working force. Do they experience a sense of self-respect, self-sufficiency, and even a degree of power that their ancestors were never able to experience? Doubtless, many working women carry a load of responsibility that is sometimes overpowering and exhausting. But for many, the experience is fulfilling. The testimony of a 58-year-old (in early 1980s) retired dining room cashier in a Beijing wheel factory speaks for countless liberated women in China.

What is the best time of life for a woman? I don’t know about other women, but for me the best time was my first job. The best day of my life was my first payday. A woman who has no work cannot control her own life. Later, when my husband died, I put all my energy into my work. If I had not had a job life would have been intolerable and I could not have borne the grief. My son is very good to me now, but I would not be happy depending on him entirely. I have my pension and my independence.\textsuperscript{57}

One of the most important aspects of achieving sexual equality for women is improving their economic status. The most efficient and effective way to do this is through education. Whether it is on the job training, night classes, literacy education or a college degree, education is a form of property for women that can be used for upward mobility. Since the first females were officially admitted to Beijing University in the late 1920s, the education of women and girls has been a top priority for women’s advocates. China has made major strides in the area of women’s literacy. The All-China Women’s Federation reported in 1999 that the illiteracy rate for adult women was 21.6 percent. Most of those women live in remote areas. The average number of years of education for women is 6.5 years. In July 1986, China implemented a Compulsory Education Law, which said that every child regardless of gender should have nine years of schools. As a result, the average number of years of education should rise with future generations of educated girls. Also reported in the year 2000 was the figure that 99.07 percent of girls in China were attending school.\textsuperscript{58}

Compare the 1916 figure of 95 percent that did not attend school with the current figure and you can quickly see that education of girls has made tremendous progress. The leading obstacle to education for girls is economics. If families can only afford to send one child to school they will send a son. Girls are often kept home because they are needed to assist with household chores. The government is addressing the problem by allocating more money for education, taking some of the financial burden off of families. Programs such as the Spring Bud Plan and Project Hope, initiated by the All-China Women’s Federation and the China Youth Development Foundation respectively, target girls in poor areas and assist them in getting or furthering their
education. At the end of 2000, it is estimated that over one million girls have been helped by the Spring Bud Plan.39

Opportunities for higher education and special training for women abound. Women have equal opportunity with men in higher education. A woman has many options when looking for further education. There are 1600 secondary vocational schools and three vocational colleges in the country, as well as junior colleges, four-year colleges and universities, and programs for Master’s and PhD degrees. In addition, there are multiple specialty schools and training centers, as well as distance learning opportunities. In 2000, 45.63 percent of women in China are in or have attended various schools or centers for adult education.50

Women in China today are also extensively involved in educational professions. Nationwide in 1992, women teachers at all levels accounted for 30-45.5 percent of the total number of teachers. One-third of the 3,000 academics at Beijing University are female, including 19 who are tutors of postgraduate candidates. In 1992 there were 20 female university presidents or vice-presidents. In addition women are very involved in areas of science and technology research. The State Council reported in 1993 that there were 8,097,000 women scientists and technicians in China. In the area of scientific research, the Chinese Academy of Sciences in 1993 had 186 women directors of research centers and 29 women academicians. Women with senior professional titles make up 40 percent of the scientists at the Chinese Academy of Medical Sciences. In 1992, 204 women were named state-level experts in their field of study.61

The advances made in women’s education over the past two hundred years allow many women and girls in China to pursue any field of study they desire. Unfortunately, at this time education is still not an option for many in remote areas. But the government of China is aware of the problems and is addressing the issue. Education can give women power to make choices and control their environments.

For more than two thousand years, women in China were in bondage to a Confucian influenced social structure that dictated they be kept in a subservient, suppressed state. Women were considered “property” to be bought and sold. The only voice a woman had was the taking of her own life to avoid being forced into an undesirable position. If a young girl survived the physical act of foot binding, the pain was constant and had to be endured for a lifetime. The foot bound woman was also physically restrained to a state of dependency.

It was the influence of outsiders that ultimately began the slow process of freeing Chinese women. Western missionaries, directly and indirectly, brought about changes in attitudes and treatment of women. The mid-19th century Taiping Rebellion had millions of followers and was the first large-scale recognition of equal rights for women. Thereafter progressive thinkers, many whom had been educated abroad, used every platform available to free Chinese women from centuries of Confucian cultural dictums.

Early in the 20th century, women became very vocal and active in their own liberation. Freedom to pursue an education, to serve in the military, and to contribute economically brought women out of the home and into the public eye. For over two thousand years half of the country’s population was suppressed and in a state of subservience, contributing to China’s lagging behind the world in modernization. Now society began to see that for China to prosper and move forward, all people had to have equal rights and opportunities.

Women in China today have the same rights and opportunities as men. Education is equally available. Women have equal rights in marriage, property and inheritance, and equal opportunity in government and politics. The government of the PRC has established women’s organizations to safeguard women’s rights and to diligently check areas of abuse. The All-China Women’s Federation, founded in April 1949, is a mass organization dedicated to further emancipation of all Chinese women.

End Notes:
1. Zhangling Wei, Status of Women: China (Bangkok: UNESCO Principal Regional Office for Asia and the Pacific, 1989), 3.
3. Wei, 6.
15. Ibid.
18. Mann and Cheng, 47-70.
Bibliography:

Beahan, Charlotte L. “In the Public Eye: Women in Early Twentieth Century China”. Historical Reflections Reflexions Historiques. no. 3 (Fall 1981): 215-238.
Cheng, Weikun. “Going Public Through Education: Female Reformers and Girls’ Schools in Late Qing Beijing”. Late Imperial China 21, no. 1 (June 2000): 107-144.
Drucker, Alison R. “The Influence of Western Women on the Anti-Footbinding Movement 1940-1911”. Historical Reflections Reflexions Historiques. no. 3 (Fall 1981): 179-199.
Ebrey, Patricia “Women in the Kinship System of the Southern Song Upper Class”. Historical Reflections Reflexions Historiques.. no. 3 (Fall 1981): 113-128.
Waltner, Ann. “Widows and Remarriage in Ming and Early Qing China.” *Historical Reflections Reflexions Historiques,* no. 3 (Fall 1981): 129-146.
Zhang Li. Interviewed by the author. 11 April 2002.

Faculty comment:

Professor and Director of the University’s Asian Studies Program, Henry Tsai, in his letter recommending the publication of Ms. Moeller’s work, wrote:

The subject matter of “Chinese Women Unbound: An Analysis of Women’s Emancipation in China” is critically important to the understanding of Chinese history. As a teacher of Chinese civilization for over three decades, I’ve come to the conclusion that suppressing half of China’s population was the key reason why China experienced so many problems in modern centuries. For over 2,000 years, Chinese women lived in a state of subservience and women were considered “property” to be bought and sold. Worse still, beginning in the eleventh century, young girls were subject to foot binding, which physically restrained them to a state of dependency. But as important as this issue is, there is a paucity of scholarship on the subject of Chinese women. Moeller is an outstanding graduating senior, has completed four semesters of Mandarin Chinese, and maintains a 3.85 grade point average. Her willingness to explore this hitherto neglected issue should be commended. She has demonstrated the kind of intellectual curiosity that *Inquiry Magazine* wants to encourage in our undergraduates.

In terms of gathering source materials at the Mullins’ Library, this subject is much more demanding and more difficult than most, and requires diligence and dexterity. Overall, this senior paper is well researched, carefully crafted, and is jammed with significant primary and secondary sources. In addition, Moeller provides a very useful bibliography and generally follows the standard Chicago format for footnote citation.
SECTION III: SCIENCE AND ENGINEERING

BIOLOGY, PLANT PATHOLOGY, MATHEMATICS, MECHANICAL ENGINEERING, AND HORTICULTURE
TRANSCRIPTIONAL CONTROL OF THE OPP OPERON
IN CLOSTRIDIUM DIFFICILE

By S. Lindsey Davis
Department of Biological Sciences

Mentor: D. Mack Ivey
Department of Biological Sciences

Abstract:

As a serious expense for both the hospital and the patient, nosocomial infections create a burden on the health care industry that is not easily overcome. Among the infections commonly contracted in the hospital environment, those associated with the Clostridium difficile bacterium account for millions of cases each year. Largely due to the nature of C. difficile infection as a response to the disruption of the normal flora of the colon caused by antibiotic activity, no completely effective treatment for this condition has been identified. It is this problem that forms the foundation for research devoted to the development of a control mechanism for the expression of the oligopeptide permease (opp) genes, which are important metabolic structural genes in the C. difficile genome. In the work reported here, a genetic construct with the ability to monitor the activity of the dual promoter known to control the expression of the opp genes was created. When this construct was used in combination with a highly expressed gene for an opp regulatory protein, we found that gene expression associated with the oppDF promoter was enhanced significantly, while oppAB expression was greatly inhibited. This pattern was observed using fructose, glucose, mannitol, and peptides as growth substrates, with the degree of induction and repression varying with the nature of the substrate. Future applications of the pUA442 construct may make it possible to determine the specific conditions that prevent the formation of structures of metabolism in C. difficile on the most basic transcriptional level, and ultimately allow for the regulation of one of the nosocomial infections that so greatly contributes to the problems of the health care industry.

Introduction:

In a recent study, 15% of observed hospital patients developed Clostridium difficile-associated nosocomial infections, resulting in 3.6 additional days of hospitalization and a 54% increase in the cost of care. From this and other related data, it was estimated that the United States spends more than $1.1 billion dollars fighting and treating these infections every year (1). As this statistical information shows, C. difficile-associated infections pose a serious problem to both the health care industry and its patients. A variety of signs and symptoms and a wide range of severity mark such infections, which may occur without symptoms, as basic diarrhea, or as any of a variety of degrees of colitis (2). However, when minor and largely asymptomatic infections go undiagnosed, serious and even fatal cases of CDAD may occur in the form of toxic megacolon and bowel perforation (3). C. difficile infections of all levels of severity are generally the result of antibiotic treatments that alter the normal flora of the colon, ultimately allowing for growth of the Gram-positive, spore-forming bacterium. Following this colonization, C. difficile produces toxins A and B, which, along with its adhesive abilities, serve as some of its most important virulence factors (4). These virulence factors prove to be quite successful, as no treatment has been consistent in combating C. difficile infections. The strong antimicrobial drugs metronidazole and vancomycin are generally used to treat C. difficile, but recurrences of infection following these therapies are common (5).

The very real implications of C. difficile infection provide a sufficient motive for studying the bacterium in order to ultimately develop a more reliable treatment, and it is the achievement of this goal through genetic control of the system responsible for C. difficile adherence and substrate utilization that forms the foundation of this research. A set of structures thought to provide these abilities has been identified in the oligopeptide permease (Opp) system of the bacterium, which in C. difficile is comprised of four subunit types designated Opp A, B, D, and F. In association with each other, these proteins of the Opp system allow for the uptake of peptides of various sizes and sequences, an activity important in C. difficile due to the preferential utilization of peptide substrates by the bacterium (6). In addition, the Opp A ligand-binding protein on the surface of the bacterial cell membrane is thought to play a significant role in the adhesive properties of the bacterium, thus acting as a factor of colonization. The implications of this information are far-reaching as control of the creation of Opp system proteins on the genetic level could possibly thwart both the adhesive and substrate utilization abilities of the bacterium, ultimately preventing C. difficile infection. The genetic locus of the Opp system has been identified and isolated by the Ivey lab, and detailed analysis has proven its composition to be relatively unique. Interestingly, the opp locus is not consistent with typical models of genetic organization in peptide permeases, instead consisting of a dual promoter controlling the oppA and oppB genes on one side and oppD and oppF oriented in the opposite direction on the other (Figure 1).
It is the activity of this unique promoter that forms the basis for the presented research, as a more complete understanding of the bacterial opp locus can provide important insight into the development of methods of control of the genes that ultimately produce the proteins of the C. difficile oligopeptide permease. If activity of the dual promoter is prevented, the opp genes it controls cannot be transcribed into messenger RNA, thus preventing translation into respective Opp proteins and the formation of the Opp system necessary for C. difficile substrate utilization and adhesion. By this chain of events, the survival mechanisms and infectious abilities of the bacterium would essentially be eliminated, along with associated pathology and disease.

On the most basic level, the achievement of these goals depends entirely on the successful production of a specific dual-reporter plasmid construct according to the methods of recombinant DNA technology and other experimental methods of molecular biology. This construct was designed so that independent genes able to transcribe two slightly different types of luciferase flank the unique C. difficile dual opp promoter region in order to provide a means of distinguishing the simultaneous expression of the oppositely oriented gene loci (Figure 2). The first of these is a firefly luciferase that emits light in the presence of luciferin (7). A measure of the light produced through the enzymatic action of luciferase can be directly correlated to the level of transcription of the luciferase gene, and thus the activity of the oppD and oppF promoter of the dual operon system. On the opposite side of the opp promoter and oriented in the opposite direction is the Renilla luciferase gene, which is able to simultaneously produce a similar but distinct luciferase when exposed to coelenterate-luciferin (8). The activity of the promoter for oppA and oppB, corresponding to the expression of this Renilla gene, can then be compared to that of the oppD and oppF genes, as indicated by levels of firefly luciferase present, concurrently and under identical conditions. Ultimately, the ability of this proposed construct to monitor promoter activation will make it possible to identify the environmental conditions that initiate and, conversely, prevent gene expression. Furthermore, understanding this bacterial stimulus-response may eventually allow for the development of a chemotherapeutic agent capable of controlling the metabolic and adhesive properties that facilitate the colonization, and thus infection, of C. difficile.

Materials and Methods:

Isolation of Plasmid DNA Construct Components

Plasmids that contain the promoter and luciferase genes required for the formation of the desired construct were isolated through the use of the Wizard™ Miniprep kit. The pUA319 plasmid was the source for the opp promoter and the gene that encodes firefly luciferase, and the pDM543 plasmid provided the Renilla luciferase gene (Figure 2). Plasmid isolations from the pUA319 and pDM543 strains were determined to be successful through analysis using agarose gel electrophoresis.

Construct Formation through Ligation-Independent Cloning

The ligation independent cloning process involves the creation of linearized segments with long nucleotide overhangs from a desired vector and insert. The bases of these overhangs are then excised by the 3′-5′ exonuclease activity of DNA polymerase until a specific recognition site is reached on each. According to their design, the products of this reaction are complementary to each other, and thus a recombinant plasmid is able to form through a simple annealing reaction and without the use of ligase. The formation and amplification of the desired linear overhang segments of each plasmid was successfully achieved through the use of the polymerase chain reaction (PCR). Agarose gel electrophoresis of each of the PCR products indicated the presence of a band of the desired size just under 5 kb for the pUA319 sample and at 1 kb for the pDM543 plasmid (Figure 3). Purification of these samples through exposure to DpnI, as well as treatment with chloroform:isoamyl alcohol and further purification using the QIAquick PCR purification kit followed. Treatment with T4 DNA polymerase according to the specific nucleotide overhangs created for each plasmid, as well as digestion with KpnI to prevent the formation of background colonies produced from any remaining template plasmid, was used to prepare the linearized samples for the annealing reaction. Annealing was performed at 22°C for 1 h. Annealed products were then used for transformation of E. coli. Subsequent plating on Luria-Bertani (LB) agar plates containing ampicillin (100 μg/ml) confirmed the success of ligation-independent cloning. Further confirmation was provided by a rough plasmid isolation using phenol and chloroform, which, through agarose gel electrophoresis, a plasmid of the desired 6kb size. The remaining cultures consistently produced 5kb bands, hinting at background growth from the original pUA319 or pDM543 plasmids. To determine the exact nature of the promising sample, the more precise plasmid purification procedure used previously for the isolation of pUA319 and pDM543 from culture was employed, this time using the QIAprep Spin Miniprep Kit.

Luciferase assays

To measure the response of the pUA442 construct to the repression of its promoters, the Dual-Luciferase, Reporter kit (9) was employed. This system is able to characterize promoter response by preparing cell cultures in such a way that the activity of the firefly and Renilla luciferase proteins produced through the transcription and translation of the genes controlled by the dual opp promoter may be quantified. Since these proteins are known to emit light as they break down specific types of luciferin, interaction with such compounds then provides a means of measurement of protein presence and activity through light emissions detected by a luminometer. Protein determination
employed the bicinchoninic acid assay, and was performed as described by the manufacturer (Pierce, Rockford, IL).

Results and Discussion:

Analysis of Purified Plasmid

To verify that the plasmid was in fact the desired construct, the purification product was first subjected to the PCR process using primers designed to amplify a portion of the desired construct that contains segments of each of the original pUA319 and pDM543 plasmids. PCR would be predicted to produce an amplified product of about 1 kb in size from the correct plasmid, but no result from any other plasmid (i.e., background). According to the electrophoresis results, gradient PCR process produced strong 1kb bands at all but the highest annealing temperature, thus providing the first proof that the plasmid isolated from the annealing reaction culture was in fact the desired construct (Figure 4A).

This confirmation was further supported through diagnostic tests using restriction endonuclease digestion. Two digests were set up according to knowledge of the location of specific restriction sites on the desired recombinant plasmid. The QIAprep-purified plasmid digested with

*Kpn I* was unaffected, as evidenced by the presence of a band just under 6 kb produced through electrophoresis (Figure 4B). This band of the same size as the isolated plasmid indicates that there were no restriction sites specific to the endonuclease, which is consistent with the expected nature of the recombinant plasmid. A double digest with the *Pst I* and *Xba I* endonucleases was also performed, with results analyzed through agarose gel electrophoresis producing bands at 4 and approximately 1.6 kb (Figure 4B). The location of unique *Pst I* and *Xba I* restriction sites on the recombinant plasmid determine that digestion with these endonucleases should produce fragments of 4146 and 1681 base pairs in size, thus providing further verification that the plasmid produced through the LIC process and isolated in QIAGEN plasmid purification was in fact the desired recombinant construct. The successfully produced plasmid construct was designated pUA442.

Test of Construct Functional Abilities:

In order to assess the capabilities of the produced construct to measure variability in *C. difficile* opp promoter activity, the effects of repressor activity on the expression of the pUA442 opp operon as represented by firefly and *Renilla* luciferase production was tested. A variable, specified as AK 


as pWSK129 is simply the kanamycin-resistant vector without the repression gene (Figure 5). Cells separately containing these plasmids plus the pUA442 construct thus provide the experimental system for the test, as the uninhibited activity of the *opp* promoter can be directly compared to that of the same structure under conditions of repression. If the pUA442 construct is functioning as designed, production of more firefly and *Renilla* luciferase proteins and thus higher luminescence measurements would be expected from the AK 


The results of the reactions set up according to the Dual-Luciferase, protocol (9) can be seen in the data presented as mean light emission values appropriately adjusted for relative protein content (Table 1). The AK 


As can be seen by comparing the variable and control in both the mid-log (3.5h) and stationary (6.5h) sample types (Figure 6), light emissions produced from the interaction of the specific luciferin recognized by each of the luciferases were consistently higher in AK 


On *ifiPTG*, samples at the same phase of growth with and without the enhancer known to increase the expression of the repressor genes present in the AK 


The expression of the repressor genes present in the AK 


expression varied according to the presence of IPTG in the AK 


However, the general trend of the AK 


It is also interesting to note that a comparison of gene expression in AK 


http://scholarworks.uark.edu/inquiry/vol4/iss1/1
Effect of Growth Substrates on opp Promoter Expression:

Following this verification of the functional abilities of the pUA442 plasmid, these capacities were put to work as the created AK, construct/repressor system was monitored in the presence of various growth substrates. Preliminary data obtained by analyzing RNA levels of opp genes suggests that expression varies in response to peptide availability (A. Richards, personal communication). Thus, cells grown in the presence of peptides or with a variety of carbohydrate substrates can serve as a model system for testing the regulation of expression of the opp operon. Due to the abilities of the created construct to act as a dual-reporter of gene expression, a comparison of luciferase production according to type under identical conditions makes it possible to determine the conditions in which both oppA and oppB, as well as the oppositely oriented oppD and oppF genes, are preferentially expressed.

In initial experimentation of this type, three carbohydrate growth substrates—fructose, glucose, and mannitol—were selected for testing and used at levels of 0.5% in LB broth media. The AK, cultures of each substrate type, as well as an LB control (in which peptides serve as the growth substrate), were further identified according to the presence or absence of the IPTG inducer. To measure levels of opp expression in these various growth conditions, cells were assayed for firefly and Renilla luciferase activity according to the previously described processes. This experimental procedure was repeated a total of three times, and results averaged to produce the mean values presented here following appropriate adjustments for relative protein content (Table 2).

Examination of these results proves interesting in many respects, the first of which focuses on the effect of the IPTG enhancer on the firefly and Renilla expression. As expected, the production and subsequent activity of Renilla luciferase is significantly lowered in those samples containing the repressor enhancer, regardless of substrate type (Figure 7A). However, consistent with observations presented in the previous luminescence experimentation, IPTG-induced production of the regulatory protein does not lead to inhibition of the portion of the dual promoter responsible for oppD and oppF production, here represented as firefly luciferase activity. Light emissions produced through such activity are, in fact, consistently higher in those samples containing the regulatory protein (Figure 7B). From these results it is possible to speculate that while enhancing expression of the portion of the promoter controlling oppD and F expression, the regulatory protein may block expression of the oppA and B portion of the dual promoter, thus allowing for the patterns of luciferase expression observed.

A more concrete conclusion made from the luminescence results verifies the hypothesis that specific substrate type has an effect on opp promoter activity, as interesting patterns of gene expression according to carbohydrate substrate or the lack thereof were identified in both the Renilla and firefly systems. Results of Renilla luminescence in enhanced samples show the highest levels of promoter activity in glucose, followed by mannitol, LB alone, and fructose, while those without IPTG were best activated in mannitol, glucose, LB, and then fructose. On the other hand, the highest levels of firefly promoter activity were found in mannitol, LB broth, glucose, and then fructose for samples both with and without the IPTG enhancer. However, due to low levels of expression of the firefly luciferase genes, it is difficult to provide a truly accurate determination of this promoter activity. From these results it can be deduced that the developed plasmid construct is able to successfully monitor variable promoter activity according to growth conditions provided by various substrates.

The successful production of a usable dual-reporter construct and confirmation of its functional and experimental capabilities has important implications for future study and, ultimately, control of the pathogenic activity of the C. difficile bacterium. As has been shown, the system can be an important tool in the analysis of the expression of the unique dual operon system of C. difficile. Based on its ability to measure transcription of the permeases of the opp locus through promoter activity, it is possible to gain a greater understanding of the specific metabolic physical conditions in which C. difficile opp genes are expressed. More extensive studies of promoter response to substrate growth conditions may ultimately establish the optimum conditions for the expression of the C. difficile opp locus, in turn facilitating the design of specific poisons that utilize the conditions determined to cause chemical deactivation and decrease gene expression. If the genes are not expressed, C. difficile will lack the ability to take up and prepare peptide substrates for use, and will be unable to survive. The utilization of this information for the development of a chemotherapeutic agent could make the prevention of the potentially serious diseases associated with C. difficile infection a reality.

References:
Table 1: Luminescence values for construct functional tests
Results of luminescence measurements for the AK1 control and AK4 repressor variable in light units. Samples with IPTG are compared to those without the enhancer in both mid-log (3.5h) and stationary (6.5h) phases.

<table>
<thead>
<tr>
<th>Samples</th>
<th>AK1 3.5h</th>
<th>AK1 3.5h</th>
<th>AK4 + IPTG 3.5h</th>
<th>AK4 + IPTG 3.5h</th>
<th>AK1 6.5h</th>
<th>AK4 6.5h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firefly 3.5h</td>
<td>0.790</td>
<td>0.426</td>
<td>1.651</td>
<td>1.094</td>
<td>0.357</td>
<td>0.333</td>
</tr>
<tr>
<td>Firefly 6.5h</td>
<td>31.05</td>
<td>28.33</td>
<td>19.76</td>
<td>15.54</td>
<td>64.14</td>
<td>28.05</td>
</tr>
</tbody>
</table>

Table 2: Luminescence values for substrate tests
Results of luminescence measurements in light units for AK4 cultured in various media both with and without the IPTG enhancer.

<table>
<thead>
<tr>
<th>Samples</th>
<th>Fructose+IPTG/Manitol+IPTG</th>
<th>Glucose+IPTG/LB+IPTG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firefly 4.131</td>
<td>4.131/9.204</td>
<td>22.78/26.57</td>
</tr>
<tr>
<td>Renilla 185.5</td>
<td>96.15/793.0</td>
<td>327.2/881.4</td>
</tr>
</tbody>
</table>

**Figure 1: Structural organization of the opp locus and promoter**

A dual promoter facilitates transcription of oppA and oppB genes oriented in one direction, and oppD and oppF in the other.
**Figure 2:** Creation of pUA442 construct from pUA319 vector and pDM543 insert. Following PCR amplification of vector and insert with desired nucleotide overhangs, activity of T4 DNA Polymerase works to excise extra bases to a specific recognition site, thus allowing for annealing of complementary ends on the desired pUA319 and pDM543 segments without the use of ligase and completing the process of ligation independent cloning.

**Figure 3:** PCR products for ligation-independent cloning. Lane 1, 1kb DNA Ladder; Lane 2, purified pUA319 PCR product for ligation independent cloning at 5kb; Lane 3, purified pDM543 PCR product for ligation independent cloning at 1kb.

**Figure 4:** Purified pUA442 PCR products at 1kb. B) Restriction endonuclease verification of pUA442 production- Lane 1: 1kb DNA Ladder; Lane 2, KpnI-digested pUA442 plasmid-no signs of restriction endonuclease cuts; Lane 3, PstI and XbaI double-digested pUA442 plasmid products at 4 and 1.6 kb.
Figure 5: Composition of AK\textsubscript{4} variable and AK\textsubscript{1} control systems. AK\textsubscript{4} contains the ampicillin-resistant pUA442 monitor construct and the kanamycin-resistant pUA328k plasmid with a gene for repression of the opp promoter, while AK\textsubscript{1} is comprised of pUA442 and the kanamycin resistant pWSK129 plasmid without an opp repressor gene.

Figure 6: Comparison of luminescence values for construct functional tests. Comparison of firefly and Renilla luminescence measurements in the AK\textsubscript{1} control and AK\textsubscript{4} repressor variable according to type and presence of the IPTG enhancer.
Faculty comment:

D. Mack Ivey, Ms. Davis's mentor, had only the most glowing things to say about her work. He wrote:

Lindsey Davis is soon to complete her senior year at the University of Arkansas, and will graduate with Honors this spring with B.S. degree in Microbiology and a minor in Spanish. Lindsey performed her thesis research in my laboratory. Lindsey is smart, energetic, curious, independent, and friendly, and has proven to be an exceptional researcher. As a junior, she devoted approximately 20 hours per week to her thesis work, while maintaining a rigorous course load (with a 4.0 GPA) as well as maintaining a busy extracurricular agenda. During the past summer, she worked full time in the emergency room of a Branson, Missouri hospital. This year, her senior year, she has continued her thesis research while taking a full course load. As both a junior and senior, she received departmental awards recognizing the highest academic achievement of any student majoring in Microbiology. She has been awarded a Silo student undergraduate research fellowship, and has received numerous academic awards, including the recent invitation to be a member of Phi Beta Kappa. She has already published an article associated with her Spanish minor.

Lindsey has studied the disease-causing mechanisms of the bacterium Clostridium difficile, a human intestinal pathogen that is associated with many cases of hospital-acquired diarrhea and colitis. She has become highly proficient in a variety of recombinant DNA techniques, and is an expert at ligation-independent cloning, which she has used to construct a luciferase dual reporter system to analyze virulence gene regulation. Her analysis of virulence gene expression has lead to the exciting and unexpected discovery of a regulatory protein with both activation and repression functions.

Lindsey is exceptional at lab work. In fact, I strongly suspect that she is exceptional at everything she does. Her experiments are always meticulously planned, and she carries them out with great dexterity. She excels at working independently, but she is not afraid to ask questions when the need arises. She handles the frustrations of the inevitable failed experiment with a very mature equanimity. She is always cheerful and positive. She is the sort of individual who brightens the space around her, wherever she is. She has willingly helped train new lab personnel, and she does more than her share of the general lab maintenance. She holds herself to a high standard, but she is not the least bit critical of others. With some delight, I have observed that her positive attitude and energy have rubbed off on others in my lab. I believe this will happen wherever she goes.

I have been at the University of Arkansas for ten years, and I have supervised an average of four to five undergraduates per year in my lab during this period. Lindsey is one of the two best students I have encountered. Her academic skills are unmatched, and her talent at research is phenomenal.

Figure 7: Expression of firefly and Renilla luciferase in cells grown on various growth substrates. Luciferase measurements were performed as described in Materials and Methods. Measurements (n=3) were taken using fresh cultures on separate days. Solid bars, mean values for samples with no IPTG added. Open bars, mean values for samples to which IPTG was added. Error bars represent the standard deviation. A) Renilla luciferase activity, and B) firefly luciferase.
THE ROLE OF NATURAL CALCIUM OXALATE CRYSTALS IN
PLANT DEFENSE AGAINST CHEWING INSECTS

By Sarah J. Doege
Department of Biological Sciences
Faculty Mentor: Kenneth L. Korth
Department of Plant Pathology

Abstract:
Calcium oxalate is the most abundant insoluble mineral found in plants and it is common among many plant families. Calcium oxalate crystals in plants can appear as irregular rectangles, spiked balls, or needles. The formidable appearance of these crystals has lead to speculation that they might serve as a form of pest control by deterring chewing insects. We utilized mutant plant lines to assess the effects of plant calcium oxalate crystals on the survival and feeding habits of chewing insects. We have taken advantage of calcium oxalate-deficient (cod) mutants of the barrel medic, Medicago truncatula. Calcium oxalate crystals accumulate in wildtype M. truncatula leaves along the vascular strands of secondary veins. Results demonstrate that beet army-worm, Spodoptera exigua, larvae that feed on M. truncatula cod mutants with reduced levels of calcium oxalate crystals grow faster and larger than insects that feed on normal wildtype plants. Pupae formed by larvae raised on cod plants are significantly larger than those raised on the wildtype plants. The results of two-way choice tests indicate that older (4th instar or later) S. exigua larvae prefer to feed on leaves lacking calcium oxalate, whereas young larvae (2nd instar or earlier) show no feeding preference. This development-specific feeding preference is perhaps due to the feeding habits of the herbivore; young larvae typically feed between secondary veins, away from areas where the calcium oxalate crystals are localized. Accumulation of RNA transcripts encoding wound-inducible gene products is normal in the cod mutants, suggesting that these plants are not altered in their ability to sense or respond to wounding by insect herbivores. Because calcium oxalate crystals seem to serve as a feeding deterrent to insects, understanding how they are made and distributed could ultimately lead to novel, environmentally sound strategies for improving insect resistance in crop plant species.

Introduction:
All plants need mechanisms of defense to limit the amount of damage by herbivores. A plant with poor defenses faces the possibility of severe foliage damage, rendering it incapable of performing an adequate amount of physiological or reproductive measures to survive. Plants defend themselves using a variety of preformed and induced chemical and/or physical barriers. The tobacco plant, for example, uses the toxic chemical nicotine as a herbivore-defense. Thorns and trichomes are examples of physical defenses that plants might utilize as protection against herbivorous insects.

Crystals of calcium oxalate are prevalent in nature and can be found in over 215 plant families (McNair, 1932). Calcium oxalate is the most common insoluble mineral in plants. Among those that produce the crystals, calcium oxalate can account for up to 3-80% of a plant's dry weight (Zindler-Frank, 1976; Libert and Franceschi, 1987). There are several hypotheses attempting to explain the role of calcium oxalate in plants - these include the regulation of bulk free-calcium, ion balance, light gathering and reflection, tissue support, and detoxification - but evidence supporting most of these hypotheses is insufficient (reviewed by Franceschi and Homer, 1980).

The placement, size, and shape of calcium oxalate crystals within some plant tissues suggest that they may play an important role in defense (Franceschi, 2001). In scanning electron microscope images, the more formidable crystals appear as multi-faceted spiked balls, or as needle-like with or without barbs. In some instances, the function of the crystals as defense mechanisms is clear, such as that of the stinging plant, Tragia ramosa (Thurston, 1976). The stinging hairs that cover its surface are conical cells, each containing a raphide (needle-shaped) crystal of calcium oxalate. Upon contact, this cell will break open, allowing the crystal to puncture the skin and administer a dermal irritant by way of grooves in the crystal. Calcium oxalate crystals in daffodil and agave are also known dermal irritants (Julian and Bowers, 1997; Salinas, et al., 2001). Many edible aroids contain the crystals in their leaves and corms, and as a result may cause swollen lips, mouth, and throat pain if prepared improperly for consumption (Bradbury and Nixon, 1998).

Research more specifically related to calcium oxalate crystals as herbivore defense reports that the dorcas gazelle, larvae of a lepidopteran species (Polytella clines), and a land snail (Eremina desertorum) all avoid feeding on certain tissues of a Negev desert lily (Pancratium sickenbergeri) that contain...
calcium oxalate raphides (Ward, et al., 1997). In other research, accumulation of calcium oxalate around the mycelia of commercially grown mushrooms has been shown to act as a deterrent of fungal gnats (Binns, 1980). To date however, there is no solid evidence that plant-derived calcium oxalate crystals serve as a defense against insect herbivores, as unrelated physical and biochemical variations in the makeup of tissues previously tested might explain observed differences in feeding behavior of herbivores.

We utilized calcium oxalate deficient (cod) mutants of the barrel medic, Medicago truncatula, to test the growth and survival of the chewing lepidopteran beet armyworm, Spodoptera exigua. M. truncatula is an increasingly important model legume used in a variety of genetic and genomic studies. This diploid relative of alfalfa (M. sativa) is useful because of its short generation time, ease of transformation, its relatively small genome size, and the increasing number of mutant lines that are available (Cook, 1999). As in alfalfa leaves, calcium oxalate typically accumulates as prismatic crystals along the secondary veins of M. truncatula leaves (Nakata and McConn, 2000).

Several M. truncatula mutant lines have been identified with reduced levels of calcium oxalate or with changes in crystal morphology or accumulation patterns (Nakata and McConn, 2000; McConn and Nakata, 2002). The mutant line cod5 completely lacks crystals, whereas cod6 has drastically reduced calcium oxalate levels. Only one genotype isolated thus far, cod4, has elevated levels of calcium oxalate as compared to the wildtype control line, A17. Five of the cod mutants show a noticeable change in crystal shape (cod1, cod2, cod3, cod6, and cod7). Large, diamond-shaped crystals are seen in cod1 and rectangular-kinked crystals are seen in cod2. Small, thin diamond-shaped crystals are found in cod3. Small, globular crystals are observed in cod6, and cod7 accumulates crystals similar to those observed in wildtype leaves, but with unusual globular protrusions at the crystal surface. Although the overall levels of total calcium in the cod mutants is not different from those in wildtype M. truncatula, the levels of calcium oxalate that accumulate can be very different. There are essentially no differences in phenotype at the whole plant level among the cod mutants, with the exception of cod4 that is stunted and chlorotic in comparison to the wildtype. All of the cod mutants described in M. truncatula are due to single-gene, non-allelic mutations (Nakata and McConn, 2000).

We report here the effects of calcium oxalate on S. exigua larvae, as measured by insect growth rates, weights, and feeding preferences on cod mutants as compared to plants with normal levels of calcium oxalate crystals.

Materials and Methods:
Insect growth experiments

S. exigua eggs were obtained from the Gast Rearing Laboratory (USDA-ARS, Starkville, MS) and allowed to hatch on artificial diet. The wildtype M. truncatula A17, which was derived from the commercial cultivar “Jemalong”, was the parental line in the mutagenesis that gave rise to the cod mutants and is used as the control treatment of all of our experiments. Plant cuttings from individual 5-week-old cod mutant lines were maintained hydroponically and placed into covered Petri dishes. At the initiation of the experiment to measure insect growth, five neonatal S. exigua larvae were placed in each dish and allowed to feed.

When sufficient plant material was available, three dishes per plant line were used. Two dishes were used for cod6 and cod4 measurements. Materials were kept in the growth chamber under a normal light: dark cycle (16:8 hr). As larvae grew to a size where it became practical to locate them among plant foliage, they were transferred to caged, intact plants maintained under the same conditions.

Larval weights were measured at intervals until pupation. Measurements ceased when larvae entered early stages of pupation, as they began to display significant weight loss associated with the process. This experiment was repeated a total of three times and in each case the larvae were allowed to pupate. Pupae were collected and length and dry weights were determined.

Preference Tests

Two independent two-choice experiments were performed to measure insect preferences for cod mutant tissue. In the first experiment three 2nd instar, or alternatively three 4th instar, larvae were placed in a Petri dish containing leaf material from both of two genotypes. Larvae of similar size were chosen and allowed to feed overnight. Weight of leaf material was measured before feeding and then after the feeding period. Dishes containing plant material were kept in larger, humidified containers to prevent desiccation of the leaf material. Ratios of leaf material consumed of each phenotype were then calculated on the basis of the fresh weight of material consumed.

In a second experiment, two-way tests were performed comparing insect preference of cod5 with A17, cod6 with A17, and cod5 with cod6, with five treatments per test. Larvae were maintained on wildtype plant material for several days before the experiment was initiated. One 3rd instar larva was placed into each dish and allowed to feed overnight. Similar amounts of leaf material were used in each treatment and the leaf material was weighed before and after feeding. The dishes were maintained as described above.

RNA blot analysis

Undamaged leaves from each genotype (A 17 and cod1 -7) were collected, frozen immediately in liquid nitrogen and then transferred to a -80°C freezer. For collection of herbivory-damaged material, S. exigua larvae (3rd or 4th instar) were caged on plants for 21 hours and damaged leaves were collected. Total RNA, 10 μg per lane, was separated via formaldehyde agarose gel electrophoresis and blotted to nylon membranes. Membranes were hybridized with radiolabeled probes for genes encoding for a GST enzyme (Ward, et al., 1997) and a COX enzyme (Ward, et al., 1997).
gel electrophoresis, transferred to nylon membrane, and hybridized with radio-labeled probes (Sambrook, et al., 1989) to measure accumulation of the wound-inducible terpene synthase transcript (tpsl). The tpsl gene was isolated from an insect-damaged M. truncatula leaf library and is strongly induced by methyljasmonate and insect damage (K. Korth, unpublished data).

Results:

Insect growth on cod plants.

Larvae of S. exigua that fed exclusively on the calcium oxalate deficient mutants cod5 and cod6 grew markedly larger than those raised on wildtype control (A17) plants. The larvae on cod5 and cod6 plants also started to pupate at least two days earlier than the controls (Figure 1). Larvae raised on the other cod mutants also showed somewhat increased growth in comparison to the controls, but still much lower than the increase exhibited by cod5 and cod6.

Larvae that fed on calcium oxalate-lacking mutants (cod5 and cod6) developed into significantly larger pupae than those that were reared on control tissue (Figure 2,A). Most of the other cod mutant-fed larvae developed into pupae that were slightly heavier than the controls. Interestingly, larvae that were fed on the cod2 mutant grew to an intermediate final pupal weight; cod2 contains a level of calcium oxalate that is intermediate between the levels in A17 and cod5. Larvae reared on the cod4 mutant, the genotype with the most calcium oxalate accumulation, showed the lowest pupal weight, although not significantly lower than controls.

Differences in pupal lengths exhibit the same trends as pupal weights on cod plants (data not shown). The difference between the pupal sizes of insects fed on A17 plants and those fed on cod5 and cod6 plants was visibly noticeable (Figure 2,B). Insect feeding preference for cod mutant tissue

We placed S. exigua larvae in closed containers with similar amounts of leaf tissue from each of two different M. truncatula lines in order to gain a measurement of feeding preference by the insects. Percentages of fresh leaf material consumed by the larvae for each set of leaves in each dish were measured after overnight feeding. From this we determined the fold-difference in the amount of food consumed in each dish. A value of 1.0 indicates that there was no preference for either genotype, whereas a value less than 1.0 would indicate that there was a preference for consumption of the A17 wildtype.

During our initial studies, we noticed that younger larvae often tend to avoid feeding on the secondary veins of A17 or cod leaf tissue. Young larvae in the 1st or 2nd instar tend to either scrape the surface of the leaf or feed between secondary veins, leaving the leaf with a webbed appearance (Figure 3,A). Young larvae perhaps had a very slight preference for cod5 leaf material over A17, whereas there was no preference exhibited between the cod6 and A17 leaf material (Figure 3,B). A dramatic preference for the cod5 and cod6 plants in comparison to A17 was clearly apparent when older 4th instar larvae were tested (Figure 3, B and C).

In a second preference test, the insects again exhibit a preference for the calcium oxalate deficient mutants over the wildtype (Figure 4). In this experiment, a single 3rd instar larva was placed in each dish with the plant material. The value of the ratio of leaf material consumed in the cod5 vs. cod6 test is near to 1.0, indicating that there is essentially no preference exhibited for either of these genotypes in this test.

Wound responses in cod mutants

AD of the cod mutants were tested for their response to insect herbivory, and virtually all of the mutants showed a strong accumulation of wound-inducible tpsl transcript in response to insect feeding (Figure 5). The wounded cod6 mutant showed a low but positive degree of transcript increase over the untreated corresponding control, however the cod6 mutant plants had previously been subjected to mild thrip damage, which might explain the transcript accumulation in the untreated control. The experiment will need to be repeated to confirm these results.

Overall, these data suggest that the cod mutants are not altered in their ability to sense and respond to insect herbivory, as indicated by defense gene induction.

Discussion:

Although it has been suggested before as a possibility (Ehrlich and Raven, 1965), there is only limited suggestive evidence that plant-derived calcium oxalate crystals serve to deter chewing insects. A specialist leafminer on American holly avoids feeding on cell layers that contain calcium oxalate crystals, and it was proposed that the crystals might serve as a mechanical barrier to feeding, although these insects were shown to feed at lower levels on crystal-containing cells in some tissues (Kimmerer and Potter, 1987). A plant low in calcium oxalate (Sinapis alba) was shown to be a more suitable aphid-host than one with higher levels (Chenopodium quinoa); the higher level of plant resistance was attributed partially to higher levels of oxalate (Liebig, 1979), but there might be other significant biochemical differences between the plant species. In contrast, one study showed that the preferred host plant for beetle (Lysathia ludoviciana [Fall] Coleoptera: Chrysomelidae) larvae was very high in calcium oxalate crystals compared to alternative hosts that were offered to the insects. The authors suggested that this host specificity might be attributable to a positive preference for the crystals by the larvae (Campbell and Clark, 1983).

We present evidence that strongly supports the hypothesis that calcium oxalate crystals are effective deterrents of chewing insect herbivory. S. exigua grew distinctly larger and pupated more quickly on plant lines with decreased levels of calcium oxalate, cod5 and cod6. Larvae reared on the other plant lines
tested also showed some increase in growth (but far from the level of cod5- and cod6-fed insects), and this may have been due to the decreased calcium oxalate in these plants or altered crystal morphology, depending on the genotype. Similar trends for larval growth were also seen in other feeding experiments using the cod mutants (data not shown). Our conclusions are strengthened by the fact that cod5 and cod6, although they are nonallelic, independent, single-gene mutations, have very similar phenotypes in terms of calcium oxalate, accumulation and in terms of how lepidopteran larvae respond to them as a food source.

We confirmed that mutant lines were not altered in their abilities to respond to wounding by insect pests. If plants were impaired in their wound-response capabilities, then greater insect growth may have been due to this and not to the absence of calcium oxalate crystals. A gene for a terpen synthase enzyme involved in the wound response pathway is expressed at low levels in undamaged A17 M. truncatula, but is highly up-regulated in insect damaged leaves. Our studies show that all of the mutant cod lines, perhaps with the exception of cod6, expressed normal levels of the terpen synthase transcript, tps1, in non-wounded and wounded leaves. This means that these plants are most likely not altered in their ability to respond to insect damage. Results for cod6 were not as clear as for the others; there was only a slight increase in expression of the gene in insect damaged leaves. Examination of tps1 transcript induction in the leaves of cod6 will be confirmed.

Not only do calcium oxalate crystals have negative effects on the growth of S exigua, but the larvae also prefer to feed on M. truncatula lacking calcium oxalate. In a preference test, 1st and 2nd instar larvae showed no preference for the cod mutants over the control A17. This may be due to the fact that younger larvae tend to feed around the secondary veins, normally avoiding the tissue that contains the calcium oxalate crystals. Because of this feeding pattern, it is likely that young larvae would have no preference for either genotype because they would not be ingesting large amounts of calcium oxalate in either case. Older larvae (4th instar and up), which usually feed on whole leaves (secondary veins and all), show much greater preference for the calcium oxalate deficient mutants. Insect preference for the cod5 and cod6 mutants was not as dramatic in a second experiment (Figure 4) as compared to an earlier measurement (Figure 3), but experimental procedure was slightly different each time. In the second experiment, only one insect per plate was used and each insect had a relatively larger amount available to be eaten. Still, preference for the cod5 and cod6 mutants was two-fold higher than that of the control plants.

Further studies in food utilization need to be applied to determine if decreased insect growth on wildtype A17 plants is due to less feeding by the insect (due to presence of an anti-feedant), or because of decreased nutritional value of the A17 plant. Identification of the genes involved in calcium oxalate crystal formation in plants might eventually allow for the targeted alteration of crystal levels, distribution and/or shape within plant tissues and contribute to the development of crops with better resistance to herbivorous insects.

Acknowledgements:

Research by S.J.D. was supported by an Arkansas SILO-SURF grant and an Adair Scholarship from the University of Arkansas Department of Plant Pathology.

Literature Cited:


Figure 1. A. Representative Spodoptera exigua larvae after feeding for an equal time period on wildtype A17 or cod5 plants, as indicated. B. S. exigua larvae grow larger and pupate earlier on plants with lower levels of calcium oxalate crystals. Neonates were caged on plants of each genotype, and larval fresh weights were determined at the times indicated. Experiments were carried out with five larvae per plant and three plants of each genotype per treatment. Plants were replaced as needed. Determination of weights was halted when larvae entered early stages of pupation.

Figure 2. Lepidopteran insects feeding on cod mutants lacking oxalate crystals develop into larger pupae. A. Dry weight of pupae reared on plants with lower levels of calcium oxalate (cod5, cod6) is significantly higher (p<0.01) than those reared on normal plants (A17). Groups not showing significant difference in dry weights are indicated by like letters on graph bars; error bars indicate standard deviation (S.D.). B. Representative pupae from insects feeding on each type of plant.
Figure 3. A. Typical feeding pattern of 1st-2nd instar larvae on M. truncatula. B. Fold-difference in relative amounts of fresh leaf tissue consumed when 2nd instar or 4th instar larvae were tested in two-way choice tests comparing cod mutants and A17. A value of 1.0 indicates that larvae consumed equal amounts of tissue from each genotype. Two larvae per plate were caged overnight with equal fresh weights of tissue. C. A typical result showing the clear insect preference for cod mutant tissue.

Figure 4. S. exigua larvae prefer cod tissue over wildtype A17 tissue, but show no preference for one mutant over another. Fold-difference in relative amounts of fresh leaf tissue consumed when 3rd instar larvae were tested individually in two-way choice tests comparing cod mutants with each other and against A17. A value of 1.0 indicates that larvae consumed equal amounts tissue from each genotype.
Figure 5. RNA blot showing normal accumulation of a wound-induced transcript in herbivore-damaged cod mutants. Leaves were fed upon by S. exigua larvae for 21 hours and accumulation of a characterized wound-inducible transcript (tps1), encoding terpene synthase, was analyzed. Transcript levels in leaves subjected to herbivory (+) were compared to untreated controls (-). The data indicate that wound signaling pathways and inducible gene expression are functional in all of the cod mutants.

Faculty Comment:

Professor Kenneth L. Korth, Ms. Doege’s mentor made the following comments about her work:

Ms. Doege has worked in my lab since the fall of 2001. She is a tremendously fast-learner, a careful and meticulous student, and pleasant co-worker in the lab. Her academic successes and the pace of her studies are notable. I have been impressed by her ability to grasp the biological significance of the experiments that she performed and other projects that are ongoing in my lab. I am even more impressed by the level of the questions she asks regarding her research, indicative of her understanding of the topics we are studying. Sarah has shown great dedication to her work, spent many long and sometimes boring hours in the lab, and has completed every task that was put before her.

With a minimal amount of guidance from myself, Sarah wrote a short proposal for the project that she has worked on. Based on that proposal, she was awarded an Arkansas SILO-State Undergraduate Research Fellowship. Funds from the fellowship were applied to the research project and also allowed Sarah to attend and present a poster at the annual meeting of the American Society of Plant Biologists in Denver, Colorado. In addition, she was awarded an Adair Scholarship to conduct research in the Department of Plant Pathology during the summer of 2002. The research project has allowed Sarah to answer some important biological questions, to learn several molecular biological techniques, to apply statistical analyses, and to get a solid introduction into lab-based biological research.

The work that Sarah completed has successfully addressed a long-held hypothesis that has, until now, been very difficult to dissect. Namely, that the calcium oxalate crystals normally found in the leaves of many plant species can serve as feeding deterrents to prevent damage by chewing insects. Scientists studying plant-insect interactions have largely overlooked the role of these common insoluble crystals in insect defense. These findings have important implications for novel mechanisms of insect control in crop plants, as the genes controlling formation of calcium oxalate might provide valuable tools for improving insect resistance. Sarah will be the first author listed on a manuscript that will soon be submitted to a peer-reviewed scientific journal.
Abstract

On the whole, political power can be very difficult to quantify. A person may be powerful due to his or her personal charm, wealth, fame, credibility, or influential connections. Political bodies do not account for these qualities when creating voting procedures; they only assign voting rules to specific positions. For example, most would say that in the United States government that a Senator is more powerful than a Representative, but less powerful than the President, without knowing any way to quantify or verify those differences.

Since the 1950's, mathematicians and political scientists have attempted to create mathematical models that partially describe an individual's power as a voting member of a committee, board, or legislative body. These models have resulted in four major "power indexes" that describe the percentage of a body's total power held by each individual member. The four most prominent power indexes are the Shapley-Shubik, Banzhaf, Johnston, and Deegan-Packel, each of which uses a different theory to calculate the probability that an individual's vote will decide whether a proposal passes or fails.

The research in this paper develops formulas to calculate the four-power indexes for legislatures that are unicameral, bicameral, unicameral with committees, and bicameral with committees. These formulas have several variables (up to ten) and have many (up to several thousand) terms for typical sizes of state legislative chambers. Using Mathematica computer software the four power indexes are computed for various legislative configurations and the indexes' behavior are studied. Then these methods are applied to the Arkansas State Government by calculating the power indexes of the Governor, Senate, House, House Committee members, and Senate Committee members. By examining the theories behind the four power indexes and available historical evidence, the paper concludes by analyzing which indexes, if any, provide the best model for the political power structure of the Arkansas State Government.

Democracy frequently requires a way to translate the various preferences of several individuals into a single group decision. The most common way of doing this is the yes-no voting system, which allows individual voters to decide between a single proposal (such as a bill or resolution) and the status quo. A yes-no voting system is defined as a set of rules that specifies exactly which collections of "yes" votes will pass a proposal [4].

Yes-no voting systems are defined in terms of coalitions, or specific collections of voters. A coalition is called winning if a proposal passes when all of its members vote 'yes' and losing if it does not meet this condition. In almost all yes-no voting systems, the winning coalitions are those that meet or exceed a quota, or the minimal number of votes for a proposal to pass. Usually, the quota is a majority, or the smallest whole number of votes that is greater than one-half of all possible votes.

The Four Power Indices:

Power indices are methods of computing the influence of an individual voter on whether a proposal passes or fails in a yes-no voting system. It is always a number between 0 and 1, and the sum of the power indices for all voters in a yes-no voting system will always be 1. The Shapley-Shubik Index was the first developed.

1. Shapley-Shubik:

In 1954, Shapley and Shubik applied Shapley's previous work regarding multi-person cooperative games to the measure of political power. Their definition of power was based on the probability that an individual player's vote will be pivotal, that is, be the qth voter when the quota is q. To understand this, suppose that all voters willing to vote for a proposal line up in some random order and vote in turn. Once q voters have voted 'yes,' the motion is declared passed and the last voter is deemed pivotal. The index assumes that all such orderings are equally probable [3]. It is computed for an individual voter A by taking the fraction of all voter orderings in which A is pivotal.
In the process of developing the formulas and applying them to various configurations, some very interesting properties of the power indices were discovered. For example, two odd-sized chambers in a pure bicameral system using simple majority votes, share Shapley-Shubik power equally. Contrary to published literature, the same is not necessarily true if one of the chambers has an even number of voters. This result has a side effect of producing some very complex combinatorial identities.

Another way of understanding the Shapley-Shubik Index is to assume that players align themselves in order of enthusiasm for a proposal, with the strongest supporter first and the strongest opponent last. There will be a number of voters “on the fence,” which the others will have to persuade to join their respective coalitions. The player who brings the coalition to winning strength is “pivotal,” and may determine, in the words of Shapley and Dubey, “how strong a law will be enacted, or how much money will actually be appropriated for some purpose, or how hard a candidate will have to campaign, etc [2].”

2. Banzhaf:

To lawyer John Banzhaf III, it was clear that [5], “voting power is not proportional to the number of votes a legislator may cast.” In a 1960s lawsuit involving the Nassau County, New York, Board of Supervisors, he was able to demonstrate that three of the six supervisors were dummies, or representatives with no actual voting power, and that the voting system of the board was therefore unconstitutional. To do so, he invented a new way to measure political power, which became known as the Banzhaf Index of Power [5].

The Banzhaf Index is based on the idea of a critical defector, a voter in a winning coalition who will cause the coalition to lose if that voter is removed from it. To calculate the Banzhaf Index for a voter \( p \), one must first count the number of coalitions from which \( p \) is a critical defector. This is called the Total Banzhaf Power of \( p \). The Banzhaf Index of \( p \) is the Total Banzhaf Power of \( p \) divided by the sum of all voters’ Total Banzhaf Power.

3. Johnston:

In 1977, R.J. Johnston, an English geographer, made some variations to the Shapley-Shubik and Banzhaf voting models and invented a new power index [6]. The Johnston Index of Power is similar to the Banzhaf Index in that it is based on critical defections. But this time, each time a voter is a critical defector from a winning coalition, it is divided by the number of critical defectors in that coalition. As before, one must calculate the Total Johnston Power of a voter \( p \), denoted TJP \( (p) \), where \( n_1, n_2, \ldots n_i \) are the number of critical defectors from each coalition:

\[
\text{TJP}(p) = \frac{1}{n_1} + \frac{1}{n_2} + \ldots + \frac{1}{n_i}
\]

To calculate the Johnston Index, \( JI \), one must divide a voter’s Total Johnston Power by the sum of all voters’ Total Johnston Power [4].

4. Deegan-Packel:

Among the basic assumptions common among the Shapley-Shubik, Banzhaf, and Johnston indices is that the power to effect change is the same as blocking power. In 1978, mathematicians John Deegan and Edward Packel suggested that these two kinds of power are actually different, and proposed an index that could measure the power to initiate changes [7]. It is based on minimal winning coalitions—those coalitions that will be losing if any one member is removed.

The Deegan-Packel Index of Power is based on three assumptions:

1) Only minimal winning coalitions are relevant when determining power.

2) Each minimal winning coalition forms with equal probability.

3) Any member of a minimal winning coalition gains the same amount of power from belonging to that coalition as all other members of the coalition [7].

To calculate the Deegan-Packel Index for a voter, one must first calculate the Total Deegan-Packel Power for that voter—the number of minimal winning coalitions of which he/she is a member. Then that number is divided by the sum of all voters’ Total Deegan-Packel Power.

Research Results:

The goals of this research had two central thrusts:

1) To develop formulas for computing each of the four power indices for four legislative models: unicameral, bicameral, unicameral with committees, and bicameral with committees.

2) Apply the formulas to the Arkansas General Assembly and its current committee structure.

Both these goals were accomplished and several significant discoveries were made along the way.

The Formulas:

Formulas were developed in each of the sixteen situations described above, and these formulas were implemented in several situations using Mathematica computer software [9]. The formulas are very complicated, having hundreds of terms and requiring calculations involving very large numbers.
example, the formula below is for the Shapley-Shubik index for one committee member in a bicameral legislature with committees. The assumption is that proposals before either chamber are first referred to a committee for a recommendation (pass or do not pass in the Arkansas legislature). If the committee recommends that the proposal be adopted, then the chamber can approve with some specified majority. If the committee does not recommend that the proposal be adopted then a larger majority is likely required for the chamber to pass the proposal. Consequently there are ten variables, $N_1$, $C_1$, $c_1$, $k_1$, $n_1$, $N_2$, $C_2$, $c_2$, $k_2$, and $n_2$, which represent the number of non-committee members, number of committee members, committee quota, chamber quota, and override quota for each house.

$$\frac{1}{(C_1 + C_2 + N_1 + N_2)!} \left( \begin{array}{c} C_1 - 1 \\ c_1 - 1 \end{array} \right)$$

$$\sum_{i=0}^{n_1-k_1} \left( \frac{N_1}{C_1 + i + k_1} \left( \sum_{n=2}^{C_2} \sum_{m=0}^{N_2} (m - n) (i + k_1 + m - 1)! \right) (C_1 + C_2 - i - k_1 - m + N_1 + N_2)! \right)$$

$$\sum_{i=0}^{n_2-k_2} \left( \frac{N_2}{C_2 + i + k_2} \left( \sum_{n=2}^{C_2} \sum_{m=0}^{N_2} (m - n) (i + k_2 + m - 1)! \right) (C_1 + C_2 - i - k_1 - m + N_1 + N_2)! \right)$$

$$\sum_{i=0}^{n_1-k_1} \left( \frac{N_1}{C_1 + i + n + 1 - 1} \left( \sum_{n=2}^{C_1} \sum_{m=0}^{N_1} (m - n) (i + n + 1 - 1)! \right) (C_1 + C_2 - m - n + N_1 + N_2)! \right)$$

$$\sum_{i=0}^{n_2-k_2} \left( \frac{N_2}{C_2 + i + n + 1 - 1} \left( \sum_{n=2}^{C_2} \sum_{m=0}^{N_2} (m - n) (i + n + 1 - 1)! \right) (C_1 + C_2 - m - n + N_1 + N_2)! \right)$$

\[\text{Power Indices for the Arkansas State Government:}\]

Both the Arkansas House of Representatives (100 members) and the Arkansas Senate (35 members) have committees that function as described above. Most House committees have 20 members [11], and most Senate committees have 7 members [12]. Executive officers with veto power that require super majorities to override (such as the U. S. President) will share power with the legislative chambers. Although the Arkansas Governor has veto power, the House and Senate can override a Governor's veto with simple majority vote [10]. Consequently, all four of the indices give the Governor zero power. The four power indices were computed for the Arkansas House and
Senate with committees of the current usual size. These are listed in the following tables and show that committee members have considerably more power than non-committee members in passing a particular bill. They also show that the power indices of the committees are affected by the existence of the other chamber. The last two columns of each table show how the relative power of each chamber is affected by the introduction of an internal committee structure.

### Properties of the Indices:

The work to establish techniques for analyzing the power structure of bicameral legislatures has yielded several important discoveries. First of all, several complex combinatorial identities have been discovered, including the complex expression for the Shapley-Shubik Index of a committee member shown above. By programming each index into Mathematica, one can create two- and three-dimensional graphs that illustrate the behavior of each power index as a function of committee size, chamber size, and quotas. Examples of such behaviors are that increasing the quota of both chambers almost always increases the power of the larger chamber, that a committee is always more powerful than a voting block of the same size, and that the four indices exhibit wildly different behaviors at extreme values.

One of the most notable results of this analysis actually contradicts an assertion originally made by Shapley and Shubik in 1954. Regarding the properties of the Shapley-Shubik Index, they stated, °∞ In pure bicameral systems using simple majority votes, each chamber gets 50% of the power (as it turns out), regardless of the relative sizes [3]. °∞ This is true if both chambers are the same size, or if both chambers have an odd number of members, which was proven by simplifying the formula for the Shapley-Shubik Index with Mathematica. The simplified expression was proven to equal 1/2 by Emeritus Professor John Duncan, using double induction on the residues of the tangent and gamma functions of a complex variable.

But examples show that two chambers do not always share power equally when one or both of them have an even number of members, such as the Arkansas State Legislature, which has a 100-member chamber with 52% of the power and a 35-member chamber with 48% of the power. This may be confirmed by visually inspecting the following two graphs. The graph on the left plots the Shapley-Shubik Index of a variable odd chamber with a majority quota against a fixed odd chamber with 49 members and quota of 25. The graph on the right plots the Shapley-Shubik Index of a variable even chamber with a majority quota against the fixed odd chamber with 49 members. Each point on the horizontal axis represents the quota of the variable chamber. Notice that the Shapley-Shubik Index of the odd chamber remains constant.
Conclusions:

This research has provided several revealing insights into the political power structure of bicameral legislatures, including the Arkansas General Assembly. The four power indices all confirm the strong influence of legislative committees, and all show that a committee is always more powerful than a voting block of the same size. The power indices also indicate some relationships that are somewhat surprising. For example, internal committees affect the relative power of two chambers. It also seems that the relative power of a committee is affected by the size and existence of another chamber.

The ability to program formulas for each power index and situation into Mathematica has lead to several meaningful discoveries about the properties of each power index. Among these properties are the effect of relative chamber size, the effect of quotas, and the overall influence of committees on a bicameral legislature.

Being able to understand the four power indices as mathematical functions lays the basis for evaluating which index provides the best measure of power for the Arkansas State Government. There are unique characteristics of each index that should be considered. For example, the Deegan-Packel Index is very consistent across the three systems—bicameral, unicameral with committees, and bicameral with committees—but does not place much power in legislative committees. The Banzhaf Index varies the most widely among the three voting situations, but gives a moderate amount of power to legislative committees. The Shapley-Shubik Index is based on fundamental concepts of game theory and distributes power in a fairly consistent manner. When complemented by historical and political analysis, these characteristics could be used to determine which, if any, of these indices are effective measures of voting power in bicameral legislatures such as the Arkansas General Assembly.

References:
Faculty Comment:

Mr. King's faculty mentor, Professor Bernard Madison had this to say about Mr. King's research:

Andrew's research area is outside anything we offer in undergraduate or graduate studies. The only related courses in mathematics are in introductory combinatorics. Consequently, Andrew began by reading through Alan Taylor's *Mathematics and Politics* (Springer-Verlag, 1995) and several research journal papers to establish a knowledge base for this research.

From the beginning, Andrew wanted to investigate mathematical measures of political power. This led to a study of yes-no voting systems, which can be considered as part of cooperative/competitive game theory. The use of game theory to study distribution of power in voting systems can be traced back to John von Neumann and Oskar Morgenstern in their 1944 classic, *Theory of Games and Economic Behavior*. Results in this area moved into the public eye with the book, movie, *A Beautiful Mind*, about the life of Nobel Laureate John Forbes Nash.

Over the past half-century, four indices of political power have been developed, often as results of legal arguments over legislative apportionment and voting rights. The four indices carry the names of their creators: Shapley-Shubik (1954), Banzhaf (1965), Johnston (1978), and Deegan-Packel (1978). The Shapley-Shubik index is best known and emerged in Andrew's work as the one receiving most attention and often making the most sense. Andrew's research focused on the application of these indexes to legislatures: unicameral, bicameral, unicameral with committees, and bicameral with committees. His main results were obtained by achieving the following two rather ambitious goals:

1. Developing expressions for computing each of the four power indices for each of the four legislative models — for committee members, non-committee members, and the full chambers.

2. Applying the four indices to the Arkansas General Assembly and its current committee structures.

Until recently, computing these indices for a state legislative body was virtually impossible because of the complexity and the large numbers involved. For example, computing the Shapley-Shubik index for the Arkansas Senate and House of Representatives requires summing several thousand terms involving ten variables and then dividing by 135, a 231-digit number! Andrew developed expressions to do this and then used *Mathematica* for the computations. The formulas are extremely complex sums of binomial coefficients and factorial expressions and represent a major accomplishment—in my view an accomplishment that is most extraordinary for an undergraduate research project.

Along the way to these two main goals, Andrew discovered results that give glimpses of the interrelationships of the indices and reveal very interesting aspects of their behavior when applied to bicameral systems. For example, Shapley and Shubik, in their 1954 paper, state, "In pure bicameral systems using simple majority votes, each chamber gets 50% of the power (as it turns out), regardless of the relative sizes. " They were referring to what is now known as the Shapley-Shubik index as a measure of power. Andrew was able to prove this (using *Mathematica* and computing residues of the tangent and gamma functions of a complex variable) if the size of each house in the bicameral system is odd and if each quota for passing a proposal is the simple majority. However, if the sizes of both chambers are not both odd (as in the Arkansas General Assembly), then the chambers need not share power equally, even when using simple majority votes. Andrew showed this by constructing examples with small numbers as well as with the Arkansas General Assembly. A side result of knowing that the above result for odd-sized chambers is a wealth of very complex combinatorial identities.

Andrew's research is significant, complex, and highly relevant. As far as I am able to discern, many of his results are original; indeed, he developed all results independently. Although mathematical power indexes cannot account for many of the aspects of political power (for example, all the indexes give the Arkansas Governor zero power in the legislative process because only simple majorities are required to override a veto), they do give insight into structural issues in legislatures. Not surprisingly, they show that committee members considering a bill have considerably more power over that bill than do non-committee members. So, Andrew's work should have wider appeal than most research in mathematics. His accomplishments far exceed what I consider normal for a strong undergraduate honors thesis, and he has achieved these accomplishments with minimal guidance from me. He has showed uncanny ability to sort through and organize some enormous counting problems, to program the resulting expressions in *Mathematica* and to interpret the results with maturity far beyond his experiences. His intuition about both mathematics and politics helped immensely.
STUDY OF NANOINDENTATION AND TIP GEOMETRY IN GaAs (100) AT ULTRA-LOW-LOADS FOR THE PATTERNING OF QUANTUM DOTS

By Robin Prince
Department of Mechanical Engineering
Faculty Mentor: Professor Ajay P. Malshe
Department of Mechanical Engineering

Abstract:

In this study, nanoindentations were produced and characterized for the future patterning of quantum dots. Nanoindentation was performed on a Si-doped (n-type) Vertical Gradient Freeze (VGF) GaAs (100) wafer with a 700 nm GaAs (100) layer grown by molecular beam epitaxy (MBE). Nanoindentation was performed with a Berkovich diamond tip, a cube corner diamond tip, and a 60° conical diamond tip. Nanoindentation of GaAs has been studied in the past, but not at extremely low loads. Previous research has been done on high load nanoindentation. The applied load in this study ranges from 400 mN all the way down to 5 mN. The motivation for achieving such low loads is to produce nanoindentations on the same size range as quantum dots (~10-100 nm in width). The smallest indentations achieved were less than 60 nm in width and less than 2 nm deep with the cube corner indenter. The geometry of the indents is characterized using atomic force microscopy (AFM).

Introduction:

The growth of particular combinations of crystalline solids, one atomic layer at a time (epitaxially), can result in the unique ability to form quantum dots. A quantum dot is an extremely small structure (~10-100 nm) that confines electrons in three dimensions, which changes the electronic and optical properties of the material. It is envisioned that these new properties will revolutionize the microelectronics and communications industries, resulting in faster, more powerful, and efficient devices. The formation of quantum dots occurs when thin semiconductor crystalline films buckle due to the stress of having lattice spacing slightly different in size from those of the substrate material. Natural growth of quantum dots results in dots of random size and location on the material substrate. Future devices will require dots with close to identical electronic and optical properties, which are dependent upon the dots possessing equivalent morphology and size, and precise spatial ordering. Thus, it is imperative to learn how to control the positioning and size of quantum dots.

Crude patterning of Indium Arsenide (InAs) quantum dots by scratching and indentation on a Gallium Arsenide (GaAs) substrate has been performed with the tip of an atomic force microscope (AFM). The indentation serves as a nucleation site for the quantum dot to grow. However, the AFM is an imaging tool, which does not allow precision control of the load placed on the surface, or the loading and unloading rate on the tip. A nanoindenter is an instrument designed to measure material properties on the nanoscale, such as hardness, Young's modulus, and plastic deformation, and is thus a more promising technique for quantum dot patterning than the AFM.

Before the quantum dot properties can be controlled with this technique, nanoindentation properties at the quantum dot scale (10-100 nm) must be better understood. Nanoindentation of GaAs has been studied in the past, but not at extremely low loads. Previous research has been done on high load (50-200 mN) and low load (200-8000 mN) nanoindentation. The loads required to produce sub-100 nm width indentations in GaAs are in a lower regime of <200 mN, which is an area that has been studied very little and has not been previously studied using a cube corner or conical indenter. The purpose of this project is to map out the nanoindentation characteristics of a Berkovich, cube corner, and conical diamond nanoindenter tip geometries at ultra-low-loads in GaAs (100).

Experimental:

The TriboIndenter® (Hysitron, Inc.) was utilized for all nanoindentation tests. The sample consists of a VGF epi-ready Si-doped GaAs (100) (supplied by American Xtal Technology (AXT Inc.)) substrate wafer with a 700nm layer of GaAs (100) grown via molecular beam epitaxy (MBE). MBE growth occurred at 580 °C with a growth rate of 0.69 monolayers per second (ML/s).

Nanoindentations were produced on the sample at room temperature with the following tips: Berkovich diamond (tip radius ~150-200 nm), 90° NorthStar cube corner diamond (tip radius ~53 nm), 60° conical diamond (tip radius <1μm). The following set of indentations were produced on the GaAs (100)
plane at each load with each tip: 5μN, 10μN, 15μN, 25μN, 50μN, 75μN, 100μN, 125μN, 150μN, 175μN, 200μN, 250μN, 300μN, 350μN, 400μN. Load, displacement, and time data was collected for each indentation. Additionally, large marker indentations were produced on the sample at 9000μN to aid in locating the smaller indents. Indentations were placed at a distance of at least ten times the indentation width apart so that strain from adjacent indentations did not affect others. All indentations were performed with a 5 second loading period, followed by a 2 second hold at the peak load and a 5 second unloading period.

Indentations were imaged with a Digital Instruments Nanoscope III atomic force microscope (AFM) at room temperature to analyze the structure of the resulting indentation, and to measure indentation width and depth.

**Results and Discussion:**

The load versus displacement curve for each indenter tip at a 400 mN peak load is shown in figure 1. It is evident from the three curves that the cube corner indenter places the highest amount of stress on the material, followed by the Berkovich then conical. This is indicated by the slope associated with the loading segment of the curve, in which a higher slope is associated with less stress. This is expected since the cube corner indenter is the sharpest of the tips, therefore supplying less area per amount of force applied. The cube corner and Berkovich tips are both 3-faceted, although the total angle of the cube corner is 90° and the Berkovich tip is 142.3°. The conical tip is in the shape of a cone with a total tip angle of 60°. However, the radius of the conical tip is large (< 1μm) compared to the Berkovich (150-200nm) and cube corner (53nm) tips. Scanning electron microscope images (SEM) of the different tip geometries are shown in figure 2.

![Figure 1 - Load vs. displacement curve for the Berkovich, cube corner, and conical tip geometries at a maximum load of 400 mN.](image1)

![Figure 2 - Nanoindentation tip geometries (supplied by Hysitron) (a) Berkovich, (b) Cube Corner, (c) 60° Conical.](image2)

Applied stress is an important consideration in the nanoindentation process, because the applied stress must be greater than the critical stress for plastic deformation to be produced. Leipner calculates that the critical stress for dislocation nucleation is ~ 6 GPa. Thus if the tip produces a stress higher than this, plastic deformation and defect nucleation is occurring. Otherwise, the tip will not make a mark in the surface at all. In order to calculate the minimum load required to produce plastic deformation, Hertzian contact theory is utilized, where load as a function of depth varies according to:

\[
\text{Load} = \frac{4}{3} E^0 \sqrt{rd^3}
\]

Where \( r \) is the tip radius, \( d \) is the depth, and \( E^0 \) is the combined tip-substrate modulus. Since the elastic modulus of diamond is much greater than that of GaAs, \( E^0 \) is taken as the modulus of GaAs (~ 99.5 GPa). Due to the radius on each tip, there exists a load at which the indentation produced will transition from spherical in shape to the geometry of the indenter. It is estimated that the spherical tip radius is all the material experiences until the tip reaches one third of the tip radius of depth into the material during indentation, therefore the width of the indentation and projected contact area can be approximated as:
w_{Berkovich}(d) = w_{Cubic}(d) = w_{Conical}(d) = 2\sqrt{(d^2 - 2rd)}
for d<r/3

Area = \pi d(2r - d) for d<r/3

Utilizing these equations, the lowest load required to produce plastic deformation, the transition load of the indent geometry from spherical to the tip shape, and the smallest indent width and depth can be calculated.

However, it was found that indentations (plastic deformation) could be produced at smaller loads than calculated above. The smallest load studied (5 μN) produced a distinct indentation with the cube corner tip, although a rough damaged area was present with the Berkovich tip at 5 μN. Additionally, the indentation shape is clearly in the shape of a triangle at 25 μN load with the cube corner tip and 75 μN load with the Berkovich tip, while at lower loads the shape of the indentation is not well defined.

Figure 3 represents the measured maximum indentation depth, maximum indentation width, and maximum pile-up height for each indent measured with the AFM for the Berkovich and cube corner indenters.

Pile-up refers to the material that is pushed up above the flat plane of the substrate. It was observed that the cube corner indenter produced a significant amount of pileup at loads ranging from 400 - 25μN, although the Berkovich tip had an insignificant amount of pile up on all indentations. Figure 4 is an example of the pile-up observed with the cube corner tip and the lack of pile-up observed with the Berkovich tip at a load of 200 μN.

### Table 1 - Calculations of indentation parameters for each tip geometry

<table>
<thead>
<tr>
<th></th>
<th>Berkovich</th>
<th>Cube Corner</th>
<th>Conical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lowest Load Required (mN)</td>
<td>83.5</td>
<td>8</td>
<td>2720</td>
</tr>
<tr>
<td>Transition Load (mN)</td>
<td>781.5</td>
<td>226</td>
<td>25500</td>
</tr>
<tr>
<td>Minimum Width (nm)</td>
<td>133</td>
<td>40</td>
<td>75</td>
</tr>
<tr>
<td>Minimum Depth (nm)</td>
<td>13.1</td>
<td>4</td>
<td>759.5</td>
</tr>
</tbody>
</table>

Figure 3 - Indentation dimensions for Berkovich and cube corner indenters versus load measured with AFM.

Figure 4 - AFM images of (a) Berkovich nanoindentation at 200 mN with no pile-up and (b) cube corner nanoindentation at 200 mN with pile-up.
A transition point from an indentation with significant pile-up to one that sinks into the plane was observed with the cube corner indenter between 25 \( \mu \)N and 15 \( \mu \)N. This is shown in Figure 5 by the AFM cross-section view of the 25 \( \mu \)N indentation and 15 \( \mu \)N indentation. The 25 \( \mu \)N indentation has significant volume of material raised above the plane, while the 15 \( \mu \)N exhibits a sinking-in effect. The varying pile-up as a function of load and tip geometry is an interesting observation because previous literature generally associates pile-up and sinking-in with the ratio of the elastic modulus to the yield stress \( (E/Y) \) and the strain-hardening properties of a material. If this were the case, one would expect to see uniformity in piling-up or sinking in for all indenter geometries at all loads, although this is not the what was observed. Further study of this effect must be conducted to determine the cause of variance associated with load and tip geometry.

At loads less than 50 \( \mu \)N the Berkovich ceased to make clear indentations and rather a raised area of material is observed at the indentation site. No indentations were expected to be visible with the conical indenter tip because the load vs. displacement curve generated during loading illustrates that the GaAs recovered completely, as indicated by the loading and unloading imposed on top of each other. No plastic deformation was observed with any of the conical indentations; however, there exists a visible area where the tip made contact with the surface. These areas are represented with a raised area of material a few nanometers in height, similar to the Berkovich indentations at loads less than 50 \( \mu \)N. It is speculated that this type of damage to the surface with no significant indentation produced could be due to the affinity between the diamond tip and the GaAs oxide layer. The stress applied at these points is not enough to produce plastic deformation, although it disrupts the oxide layer and draws the oxide up with the tip as it leaves the surface.

**Conclusion:**

The geometry of the nanoindenter tip has a significant effect on the type of indentation produced in GaAs (100). The conical indentation does not produce indents at all, but rather a damaged area in the surface at loads less than 400 mN. The Berkovich and cube corner indenters make 3-faceted indentations up to 50 mN and 25 mN loads respectively. Although at loads less than 50 mN with the Berkovich tip there is a raised area of damage at the indentation site similar to what is observed in the conical indentations. The cube corner continues to induce...
plastic deformation and leave indentations at 15 mN, 10 mN, and 5 mN, although the indenters are not as reflective of the tip geometry and no material pile-up is observed around the indentations, while significant pile-up is observed in cube corner indentations at loads higher than 25 mN. At high loads with the Berkovich tip, no significant amount of material pile-up is observed around the indentation. Quantum dots are on the order of ~10-100nm in width. Therefore, it is hypothesized that the cube corner indentations at 15 mN and less will be the best size and shape for quantum dot growth, because there is very little pileup and no sharp edges. These indents produced at <15 mN were less than 58 nm in width and 2.5 nm in depth. In order to grow quantum dots at the nanoindentation sites, the indented GaAs wafer would be placed into the MBE chamber and brought up to a temperature around 600°C before growth of the InAs layer. It is hypothesized that the shape of the indentations will change as an effect of annealing at the MBE growth conditions, therefore future work will include annealing of the samples and re-evaluating the nanoindentation properties.

Acknowledgements:

We would like to acknowledge that this research was partly sponsored by the National Science Foundation (NSF) and Research Experience for Undergraduates (REU-NSF Grant EEC-0097714) and Integrated Graduate and Research Traineeship (IGERT-NSF Grant 9972820) programs.

References:

Faculty comment:

Professor Ajay Malshe, Ms. Prince’s mentor, feels that her research has important implications for further scientific study. In his letter of nomination to the publication board, he wrote:

Ms. Robin Prince is the brightest undergraduate student that I have guided in the last seven years of my faculty career. Robin made her excellent academic and dynamic professional impression because of her passion for scientific curiosity, cutting edge technology and ability to function in a team while maintaining her identity. She has been working on “studying nanomechanics using nanoindentation at tip-GaAs interface” in my research group for the past two years. Her participation is funded through an NSF-Research Experience for Undergraduates (REU) program. The project is an interdisciplinary effort among Physics and Mechanical Engineering departments in the field of Nano Integrated Micro Systems.

The investigation on characterizing and controlling damage (i.e. stress, strain, lattice defects) in gallium-arsenide (GaAs) caused by nano-scale mechanical perturbation for the realization of patterned quantum dot arrays is the area of Ms. Robin Prince’s broader scientific interest. Specially, the aim of Ms. Prince’s project and manuscript was to investigate the role of nano indentation tool tip geometry on strain and morphology of the nano imprint patterning of quantum structures. Precision spatial ordering or ‘printing’ of quantum dots by the injection of highly-localized damage sites in 111-V semiconductor surfaces is an area of immense scientific, engineering, and commercial applications interest.

Her approach is to study patterned perturbations in GaAs caused by nanoindentation. The study of nano-scale perturbations in GaAs will give data and clues to the growth mechanisms of quantum dots as a function of the perturbation site’s nanomechanical properties. If this research is successful, it would establish processes for the selective growth of quantum structures and would lead to viable and inexpensive nanomanufacturing processes for fabricating arrays of identical quantum dots. Ms. Prince’s research addresses the following research and education themes: (1) Manufacturing Processes at the Nanoscale, (2) Nanoscale Structures, Novel Phenomena, and Quantum Control, and (3) Nanoscale Devices and System Architecture.
RESPONSE OF BLACKBERRY CULTIVARS TO NEMATODE TRANSMISSION OF TOBACCO RINGSPOT VIRUS

By Alisha Sanny
Department of Horticulture

Faculty Mentors: Professors John R. Clark and Rose Gergerich
Departments of Horticulture and Plant Pathology, respectively

Abstract:
A study was conducted on eight cultivars of blackberry (‘Apache’, ‘Arapaho’, ‘Chester’, ‘Chickasaw’, ‘Kiowa’, ‘Navaho’, ‘Shawnee’, and ‘Triple Crown’), of which four plants of each were previously determined in the fall of 2001 to have root, but not leaf, infection with Tobacco ringspot virus (TRSV). The objective of our study was to determine virus effects on plant vigor, and the spread of virus infection in the plants. Eight plants of each cultivar, four infected and four free of infection, were grown in pots on a gravel pad for the 2002 growing season, and samples of primocane and floricanes were taken to determine if TRSV had moved to the above-ground portion of the plants. TRSV infection was determined by ELISA tests. At the end of the growing season (October), the plants were harvested and dry weights determined for floricanes, primocanes, and roots to determine virus effects on plant vigor. In all plants that had been shown to have root TRSV infection, the virus was shown to have moved into the top portion of the plants as evidenced by positive ELISA tests on primocane and floricanes. The objective of our study was to determine virus effects on plant vigor, and the spread of virus infection in the plants. Eight plants of each cultivar, four infected and four free of infection, were grown in pots on a gravel pad for the 2002 growing season, and samples of primocane and floricanes were taken to determine if TRSV had moved to the above-ground portion of the plants. TRSV infection was determined by ELISA tests. At the end of the growing season (October), the plants were harvested and dry weights determined for floricanes, primocanes, and roots to determine virus effects on plant vigor. In all plants that had been shown to have root TRSV infection, the virus was shown to have moved into the top portion of the plants as evidenced by positive ELISA tests on primocane and floricanes. The objective of our study was to determine virus effects on plant vigor, and the spread of virus infection in the plants. Eight plants of each cultivar, four infected and four free of infection, were grown in pots on a gravel pad for the 2002 growing season, and samples of primocane and floricanes were taken to determine if TRSV had moved to the above-ground portion of the plants. TRSV infection was determined by ELISA tests. At the end of the growing season (October), the plants were harvested and dry weights determined for floricanes, primocanes, and roots to determine virus effects on plant vigor. In all plants that had been shown to have root TRSV infection, the virus was shown to have moved into the top portion of the plants as evidenced by positive ELISA tests on primocane and floricanes. The objective of our study was to determine virus effects on plant vigor, and the spread of virus infection in the plants. Eight plants of each cultivar, four infected and four free of infection, were grown in pots on a gravel pad for the 2002 growing season, and samples of primocane and floricanes were taken to determine if TRSV had moved to the above-ground portion of the plants. TRSV infection was determined by ELISA tests. At the end of the growing season (October), the plants were harvested and dry weights determined for floricanes, primocanes, and roots to determine virus effects on plant vigor. In all plants that had been shown to have root TRSV infection, the virus was shown to have moved into the top portion of the plants as evidenced by positive ELISA tests on primocane and floricanes. The objective of our study was to determine virus effects on plant vigor, and the spread of virus infection in the plants. Eight plants of each cultivar, four infected and four free of infection, were grown in pots on a gravel pad for the 2002 growing season, and samples of primocane and floricanes were taken to determine if TRSV had moved to the above-ground portion of the plants. TRSV infection was determined by ELISA tests. At the end of the growing season (October), the plants were harvested and dry weights determined for floricanes, primocanes, and roots to determine virus effects on plant vigor. In all plants that had been shown to have root TRSV infection, the virus was shown to have moved into the top portion of the plants as evidenced by positive ELISA tests on primocane and floricanes. The objective of our study was to determine virus effects on plant vigor, and the spread of virus infection in the plants. Eight plants of each cultivar, four infected and four free of infection, were grown in pots on a gravel pad for the 2002 growing season, and samples of primocane and floricanes were taken to determine if TRSV had moved to the above-ground portion of the plants. TRSV infection was determined by ELISA tests. At the end of the growing season (October), the plants were harvested and dry weights determined for floricanes, primocanes, and roots to determine virus effects on plant vigor. In all plants that had been shown to have root TRSV infection, the virus was shown to have moved into the top portion of the plants as evidenced by positive ELISA tests on primocane and floricanes.

Introduction:
There have been 26 virus or virus-like diseases reported for Rubus crops (blackberries and raspberries) in the world (Jones, 1986). Viruses cause more damage in the black and purple raspberries and less damage in red raspberries and blackberries (Crandall, 1995). Many viruses that infect blackberry do not produce distinctive symptoms, and reports of virus effects on blackberries are very limited. A recent study on the impact of Raspberry bushy dwarf virus (RBDV) on ‘Marion’ blackberry in Oregon showed that there was no virus effect on cane number or length in a two-year period, but that there was a significant yield reduction (50%) in RBDV infected plants, along with reduced berry weight (40%) and drupelet number per berry (39%) (Strik and Martin, 2002). Infected plants also showed visual symptoms, including chlorosis, vein clearing, silver discoloration, and malformed, small fruit. Newly infected plants did not display such distinct symptoms.

There are approximately 200-400 acres of blackberries grown and marketed locally throughout Arkansas. Eleven licensed Arkansas nurseries and 29 licensed nurseries in other states and countries propagate University of Arkansas patented cultivars, as well as other cultivars, for national and international markets (Troxell, 2001). The presence of virus symptoms in nurseries and commercial blackberry fields in Arkansas has been a recent cause for concern. A field survey was conducted in Arkansas of blackberry nurseries for TRSV, RBDV, and Impatiens necrotic spot virus (INSV) in 2002. All three viruses were found, but TRSV was found first and was most prevalent (Rose Gergerich, unpublished).

Leaves of blackberry and dewberry plants in North Carolina infected with TRSV showed faint to severe ringspots, mottling, mosaic, stuntling, leaf distortion, and yellow line patterns (Rush and Gooding, 1970). However, they usually did not have symptoms on each cane. Virus symptoms on blackberry plants from Arkansas showed chlorosis, oak-leaf patterns, and mosaic (Troxell, 2001; Fig. 1).

Guzman, et al. (2002) more recently conducted a virus survey in North Carolina, South Carolina, and Virginia in 2001-2002. In North Carolina, TRSV was found in 57% (257/451) of the symptomatic plants tested. TRSV was identified in 33% (21/62) of the samples from South Carolina, but TRSV was not identified in any of the plantings tested in Virginia. TRSV was detected in ‘Apache’ and ‘Arapaho’ most frequently (>50%). It was also detected in ‘Chester’, ‘Chickasaw’, ‘Lochness’, and ‘Rosborough’.

TRSV is a nepovirus that was first reported in Nicotiana tabacum by Fromme et al. (1927). The first report of TRSV in blackberry was published in 1965 from North Carolina (Stace-Smith, 1987). Rush and Gooding (1970) isolated TRSV from
Rubus allegheniensis, R. argutus, R. flagellaris, and an unidentified Rubus species in North Carolina. TRSV has a large host range, including both herbaceous and woody plants (Stace-Smith and Hansen, 1974). It occurs throughout North America, especially in the southeastern United States (Rush and Gooding, 1970).

TRSV virions are isometric, not enveloped, and 25-29 nm in diameter. The genome consists of two single strands of linear RNA, both of which are needed for infection (Brunt et al., 1996). The primary spread of TRSV in the field is by the dagger nematode, Xiphinema americanum. The virus does not multiply in the vector and is lost once the nematode molts. It can also be transmitted mechanically, by infected nursery stock, and by pollen (Brunt et al., 1996).

In the most extensive study of TRSV on blackberry, Troxell (2001) conducted nematode transmission experiments on eight blackberry cultivars (Apache, Arapaho, Chester, Chickasaw, Kiowa, Navaho, Shawnee, and Triple Crown). She infected tissue-cultured plants with TRSV using X. americanum transmission, and found all cultivars were susceptible to this virus as determined by sampling roots of exposed plants and testing for TRSV using Protein-A ELISA tests. ELISA tests revealed that TRSV was not present in the leaves of aboveground portions of these plants during the first growing season following nematode transmission of the virus. Symptoms were seen on primocane leaves of infected plants, but these were mild and transient.

Our study was initiated to further evaluate the virus effects on virus-infected or non-infected plants used in 2001 by Troxell (2001). Specifically, we wanted to determine the impact, if any, of TRSV infection in the second year following nematode transmission on blackberry plant vigor and to find if the virus could be found in aboveground plant portions based on ELISA tests.

Materials and Methods:

In March 2002, four non-infected and four infected plants of each cultivar used in an earlier study (Troxell, 2001) were chosen for this study. The plants had been grown in 3-L plastic pots the previous season, and had been mulched with sawdust mulch over the winter to protect them from winter injury. In March 2002, the plants were removed from the mulch and pruned, leaving two 0.8-m floricanes. Floricanes were staked using 1-m plastic stakes. The plants were moved to a gravel pad at the Arkansas Agricultural Research and Extension Center, Fayetteville, and four plants (replications) of each cultivar/virus infection status combination were arranged in a randomized block design. Each pot was set up in 12.2-m rows on 1.5-m squares (Fig. 2). Starting in May, pots were fertilized every two weeks with one tablespoon Osmocote (19-6-12) (until July) and irrigated overhead twice daily for one hour (until September). Plants were observed from May to November for virus symptoms.

In July 2002, Protein-A ELISA, as described by Edwards and Cooper (1985), was used for the detection of TRSV in the floricanes and primocanes of all plants. Leaf extracts were prepared by grinding floricanes and primocanes leaves with a sap extractor (Erich Pollakhe Co., Wennentgen, West Germany). The sap was diluted 1:10 (v:v) in extraction buffer (10.3 mM Na_2SO_4, 0.5 mM polyvinylpyrrolidone M.W. 40,000, PVP-40), 0.2% powdered egg albumin, 2% Tween-20 in PBS-T [137.8 mM NaCl, 1.47 mM KHPO_4, 8.1 mM Na,HPO_4, 2.68 mM KCl, 0.05% Tween-20, pH 7.4]. Immunol 1 flat-bottom microtiter plates (Dynex Technologies, Inc., Chantilly, Va) were coated with Protein-A (Sigma Chemical Co., St. Louis, Mo.) at 1 ug/mL in coating buffer (15 mM Na_2CO_3, 34.88 mM NaHCO_3, 3.08 mM Na_2SO_4, pH 9.6). Plates were then coated with polyclonal antiserum to TRSV (from University of Arkansas plant virus antiserum collection) that had been diluted 1:1000 with PBS-T. After addition of the diluted leaf extract to duplicate wells of the plate, TRSV antiserum diluted 1:1000 in PBS-T was added to the plates followed by Protein-A alkaline phosphatase conjugate (Sigma) at 1 mg/mL in PBS-T. Finally, nitrophenyl phosphate at 1 mg/mL (Sigma) in substrate buffer (0.39 mM MgCl_2, 3.84 mM Na_2PO_4, 12.1% diethanolamine; pH 9.8) was added to the wells. Between all steps, plates were washed three times with PBS-T. All reagents were used at 100uL/well, and incubations were at 24°C for 2 h, except for the incubation after antigen addition, which was at 4°C overnight. Absorbance at 405 nm was determined with a 7520 Microplate Reader (Cambridge Technologies, Inc., Watertown, Mass.).

Plants from three replications were defoliated in November 2002 by hand, and the floricanes and primocanes cut at the crown level (soil surface) and dried separately in paper bags for 4 O days at 65°C. After cans were dry, they were weighed and discarded. Roots were washed, dried in the same manner, and weighed. The fourth replication was kept for further observation. Dry weight data were analyzed as a two-factor randomized complete block by JMP (JMP, version 4.0. SAS Institute, Inc. Cary, N.C., 1989-2000).

Results and Discussion:

All plants of all cultivars that had tested positive in ELISA tests in root samples taken in 2001 tested positive for TRSV in primocane and florican leaves in July 2002. The ELISA tests also demonstrated that the non-infected control plants continued to be virus free. The finding that virus was present in the leaves indicates that the virus moved from the roots to the aboveground portion of the plants in the second year after nematode transmission. No virus symptoms were observed on the leaves during the study on infected or non-infected plants.

For the dry weight data, the data analysis of variance indicated no significant (P=0.05) interaction of infection status and cultivar for the variables measured. Additionally, the main effect of virus infection status was not different for any variables,
indicating no virus effect on plant vigor. The dry weight means were similar for floricanes, primocanes, and roots (Table 1). Although no significant differences were found between infected and non-infected blackberry plants in this study it is possible that the primocanes which are now virus infected will show symptoms of virus infection when they develop as floricanes next year. Cultivars averaged over plant infection status were significantly different for dry weight of primocanes, floricanes, and roots (data not shown). This finding was not important for this investigation since our effort was to identify virus effects, not cultivar vigor differences.

The most noteworthy finding of our study was the second-year presence of TRSV in leaf samples. This has a number of implications for management of blackberry virus diseases. First, it is important for nursery growers and regulatory agencies because plants that become infected in the field by nematode transmission may be carrying virus in their roots (the portion of the plant often used for propagation) but test negative for virus in their leaves in ELISA tests and appear healthy based on the lack of leaf symptoms. Second, the dynamics of TRSV movement in the plant are clearer from our results. The delay in virus movement from the roots to the aboveground portion of the plant was longer than expected. This area needs further study to determine if leaf symptoms develop on the floricanes which became infected with TRSV during the last year.

Field-grown, mature blackberry plants that express virus symptoms often grow well and bear abundant, good quality fruit. This has raised the question “What is the effect of viruses such as TRSV on blackberries”. Our data on second-year plants indicates no virus effect on total plant growth. However, the virus may have greater effects the third year as it spreads further in the plant. Also, our plants did not bear fruit, thus we were unable to determine if TRSV-infected plants produced fruit that was malformed, crumbly, or otherwise possibly affected by virus infection. Additionally, if blackberry plants become infected with more than one virus, the combined effects of these viruses often produce severe disease symptoms.

The concerns of blackberry nursery stock producers are somewhat different from those of blackberry fruit producers. Many states and countries have regulatory agencies that restrict the movement of plants that are not certified as virus-tested. Blackberry plants expressing symptoms of virus infection would be denied entrance by such regulatory agencies. Based on the results of the research reported here, the absence of virus symptoms in blackberry leaves should not be used to determine whether blackberry plants are infected with TRSV.

Literature Cited:
Table 1. Main effect means (dry weight in grams) for non-infected vs. infected blackberry cultivars.

<table>
<thead>
<tr>
<th>Virus status</th>
<th>Primocanes</th>
<th>Floricanes</th>
<th>Roots</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infected</td>
<td>160.1</td>
<td>31.3</td>
<td>590.5</td>
</tr>
<tr>
<td>Non-infected</td>
<td>156.5</td>
<td>26.5</td>
<td>640.3</td>
</tr>
</tbody>
</table>

Significance: F-test, P=0.05.

Faculty Comment:

Professor John Clark, one of Ms. Sanny’s faculty mentors, made the following remarks about her work:

I have served as Ms. Sanny’s undergraduate advisor, and co-advisor for her research project. Working with her has been a true joy! She came to the University as a freshman Chancellor’s Scholar, and has excelled academically. I cannot overstate the extraordinary academic performance by her while enrolled here. She also has been an exemplary undergraduate researcher. She was able to undertake this study and complete it very well, and prepare the manuscript in a very timely manner.

The research study, co-advised with Dr. Rose Gergerich in Plant Pathology, dealt with investigating the movement or within-plant spread of Tobacco ringspot virus in a number of blackberry cultivars. This virus is of prime importance in Arkansas and other southern states as it is being found in commercial blackberry fields commonly. However, little is known of its movement in the plant or the overall effect of the virus on the plant. Her study showed that after infection, the virus moved into the above-ground parts of the plants the following year. She also found that from an overall plant vigor standpoint that the virus did not affect the plant’s growth in the second year after infection. These are both findings that have not been reported in blackberry before. This information will be useful for further research on this topic, and also will contribute to better inspection by Arkansas State Plant Board of blackberry nurseries that propagate these cultivars.

Fig. 2. Blackberry plants in field trials, mid-summer, at the Arkansas Agricultural Research and Extension Center, Fayetteville.