

Journal of
Undergraduate Research

**UMBC
REVIEW
2012**

vol. 13





© COPYRIGHT 2012

University of Maryland,
Baltimore County
All rights reserved.

editors: Melanie Dell
Esther Gross

designers: Brendan Lipton
Faryal Khalid

cover photograph by: Brendan Lipton



Journal
of
Undergraduate
Research

UMBC
REVIEW
2012

vol 13

table of contents

- 10 Michael Curtis
Mathematics and Statistics
- 22 Samuel Garrett
Music
- 36 Danielle Viens-Payne
Modern Languages, Linguistics,
and Intercultural Communication
- 64 John Winder
Art History and Computer Science
- 90 Xuan Ge, David Stonko
Biology and Mathematics
- 114 Kelly Jordan
History
- 128 Asmara Qamar
Biology
- 146 Christina Briscoe
Interdisciplinary Studies
- 182 Simbarashe Marufu
Physics
- 198 Joshua Hunter
Political Science
- 

EDITORS' INTRODUCTION

THIS IS THE thirteenth edition of the *UMBC Review: Journal of Undergraduate Research*. Featuring a wide variety of articles, the *UMBC Review* is a multi-disciplinary journal that highlights the original work of UMBC's undergraduate students. This journal, published annually, is run by student editors and designers with the help of faculty advisors. Each article that is published in the *UMBC Review* is selected through a rigorous and competitive faculty- and peer-reviewed process.

The *UMBC Review* provides an opportunity for undergraduate students from any discipline to have their research considered for publication. As undergraduates, many students do not have the chance to get their research published and to receive recognition for their work. The rewarding opportunity offered by the *UMBC Review* encourages undergraduates to pursue research that interests and inspires them and is relevant in today's academic setting.

Furthermore, the *UMBC Review* allows student authors to take an active role in the publication of their research, starting with the writing of their manuscripts. If their papers are selected, the authors work closely not only with their advisors, but also with us throughout the entire revision process. Through the faculty-review process, we encourage authors to connect with other faculty members within their disciplines. By facilitating and fostering communication between students and professors, the *UMBC Review* prepares students for the collaborative aspect of research.

UMBC has been ranked the nation's #1 "Up and Coming" university in 2009, 2010, and 2011 by *U.S. News & World Report* and was featured in November 2011 on CBS News' *60 Minutes*. Among the many qualities that make UMBC stand out is its exceptional commitment to its undergraduate students. One of the ways that UMBC illustrates this dedication is by providing an undergraduate research journal for the benefit of its scholars. For instance, the existence of the *UMBC Review* on campus fosters awareness among students of the importance of sharing their research ideas with the greater academic community. We feel that the dissemination of their peers' research is an integral part of encouraging students to engage in research at an early, undergraduate level.

We are proud to have the privilege to showcase a sample of the outstanding undergraduate research that has been conducted on this campus. The articles in this year's journal have inspired us.

We hope that they capture your interest and increase your knowledge on the following topics:

Inspired by the “Netflix Million Dollar Challenge,” **Michael Curtis** applies mathematical and statistical techniques, including a collaborative filtering model, to provide a method of predicting users’ future movie ratings based on their previous choices.

In his research from the discipline of music, **Samuel Garrett** conducts spectral and spatial analyses of “Pilentze Pee” from *Le Mystère des Voix Bulgares* to uncover the qualities that make this piece aesthetically powerful and fascinating.

From the Department of Modern Languages, Linguistics, and Intercultural Communications, **Danielle Viens-Payne** presents ways to help students with learning disabilities succeed in foreign language classes in secondary schools and institutions of higher education.

John Winder combines the different disciplines of art and computer science in his research paper, which provides a discussion on the history and use of chiaroscuro (shadows and shading) in European oil paintings from 1500 to 1900 and shows how the algorithm he created can facilitate the exploration of light and dark in these paintings.

In their interdisciplinary article, **Xuan Ge** and **David Stonko** combine mathematical modeling and biology to analyze border cell migration in fruit flies.

From the discipline of history, **Kelly Jordan** analyzes the portrayal of women in 1950s sitcoms to show how this representation provided families with a model of idealized female roles in the insecure, early Cold War American environment.

Asmara Qamar presents her biological research on the characterization of *Escherichia coli* mutant strains that are resistant to the antibiotic cethromycin.

From the field of interdisciplinary studies, **Christina Briscoe** discusses the research she conducted in Praia Grande, Brazil, particularly focusing on the varying experiences of young mothers who live in this marginalized area.

Simbarashe Marufu’s physics article examines surface modification using self-assembled monolayers in order to facilitate atomic layer deposition.

In his political science article, **Joshua Hunter** explores the identity, role, and function of the North Atlantic Treaty Organization (NATO) in order to explain the reasons behind NATO's existence beyond the Cold War and to prove the relevance of this alliance today.

We were very impressed by the intellectual creativity of the authors who submitted articles to this year's edition of the *UMBC Review*. We extend our utmost thanks to these students for their dedication to their research. We would also like to recognize the individuals below who helped to make the *UMBC Review* possible.

ACKNOWLEDGMENTS

WE WOULD FIRST like to thank the authors' faculty advisors, who provided countless hours of support for their advisees. Their guidance ranged from helping these students choose the right research topic to assisting them in the writing process:

Dr. Bambi Chapin (*Department of Sociology and Anthropology*)

Dr. Linda Dusman (*Department of Music*)

Dr. Theodosia Gougousi (*Department of Physics*)

Dr. Preminda Jacob (*Department of Visual Arts*)

Dr. Nagaraj Neerchal (*Department of Mathematics and Statistics*)

Dr. Ana Oskoz (*Department of Modern Languages, Linguistics, and Intercultural Communication*)

Dr. Meredith Oyen (*Department of History*)

Dr. Bradford Percy (*Department of Mathematics and Statistics*)

Dr. Brigid Starkey (*Department of Political Science*)

Dr. Michelle Starz-Gaiano (*Department of Biological Sciences*)

Dr. Janice Zengel (*Department of Biological Sciences*)

We are very appreciative of the professors who served as faculty reviewers. These professors from several universities willingly volunteered their time to provide constructive criticism on the authors' papers. For confidentiality purposes, their names will remain anonymous.

We extend our thanks to the following proofreaders of the *UMBC Review*, especially Atheeth Hiremath, last year's *UMBC Review* editor, who provided additional help editing the STEM papers:

Laura Bartock	Janet McGlynn
Linda Brown	Tom McGlynn
Devon Fick	Diana Miller
Ben Goldberg	Nancy Miller
Pam Hawley	Jess Myers
Sally Helms	Anne Roland
Atheeth Hiremath	Laila Shishineh
Jarrett Kealey	Beth Wells
Stephanie Lazarus	

We would also like to convey our sincere appreciation to the people who are exceptionally devoted to providing funding and support for UMBC's undergraduate research and the *UMBC Review*:

Dr. Freeman Hrabowski
The Office of Undergraduate Education
The President's Office
The Provost's Office

Finally, we would be remiss if we did not express our gratitude and thanks to the two faculty advisors of the *UMBC Review*. They were always available to guide us through any step during the creation of this year's journal:

Dr. Raphael Falco (*Editing*)
Guenet Abraham (*Designing*)

We hope you enjoy the 2012 edition of the *UMBC Review*.

EDITORS

Melanie Dell
Esther Gross

A COLLABORATIVE FILTERING MODEL Statistical Properties of Alternating Least Squares

michael curtis

Michael Curtis has attended UMBC for four years, pursuing a B.S. in Computer Science with a concentration in Game Design. After participating in UMBC's High Performance Computing Research Experience for Undergraduates (REU) directed by Dr. Matthias Gobbert and Dr. Nagaraj Neerchal, he decided to also pursue a B.A. in Mathematics. He will graduate in May 2012 and hopefully continue studying computer science in graduate school. While at UMBC, Michael was a participant in the Meyerhoff Scholars Program, a member of the Game Development Club, and a DJ at WMBC. He would like to thank Lamont Toliver and the Meyerhoff Scholars Program staff, the Department of Mathematics and Statistics, and Dr. Neerchal and Dr. Gobbert for encouraging him to challenge himself with a major in mathematics.

This research began as a University of Maryland, Baltimore County REU site project in the summer of 2010. With the help of Dr. Robert Bell, a member of the winning team of the Netflix prize, our goal was to build a recommender model and then parallelize the algorithms. My interest in parallel algorithms made the project a perfect fit. Because I was only a computer science major, I did not understand the complete model, but I had a major role in its code. This past year, Dr. Nagaraj Neerchal, Andrew Raim, and I decided to take a closer look at one of the algorithms, specifically Alternating Least Squares (ALS). We put it through many error-checking tests and finally felt comfortable that we had a model that represented the problem of recommending very well. This helped me to gain an understanding of what I helped create, a model which I believe will ease the transition from serial to parallel code. This paper is our close examination of the ALS algorithm handling sparse data.

INTRODUCTION

RECOMMENDER SYSTEMS ARE emerging as important tools for improving customer satisfaction by mathematically predicting user preferences. Several major corporations, including Amazon.com and Pandora, use these types of systems to suggest additional options based on current or recent purchases. Netflix uses a recommender system to provide its customers with suggestions for movies that they may like, based on their previous ratings. The objective of the recommender system is to obtain a model that predicts future ratings a user might give for a specific movie. One such model is known as a collaborative filtering model. When applied to movie selection, the model includes the average movie rating (μ), the rating bias of the user (b), the overall popularity of the movie (a), and the interaction between user preferences (p) and movie characteristics (q). The collaborative filtering model makes use of the sparse data collected from users who have watched movies and provided ratings. With the data, the model obtains predicted ratings for movies the users have not yet watched. Collaborative filtering models have had growing interest recently due to the Netflix million-dollar challenge to improve its existing algorithm for recommending movies. Through the contest, Netflix hoped to enhance the method with which they recommend movies to their users. This method would be based on prior movies the users have rated. The goal of the contest was to predict, as accurately as possible, what any particular user would rate a movie.

Developing the next-generation filtering model for Netflix could be interpreted as a standard regression problem: fitting a model that relates the characteristics of movies with those preferred by users to develop an estimated rating of the user. However, a closer look into the problem reveals that it is more challenging. The Netflix data does not explicitly contain any information about movie or user characteristics, except for ratings that users have given to movies. Since movie characteristics are not readily available as regressors, the model cannot be treated as a regression model. The ratings data are also sparse, e.g. some users have rated many movies, others hardly any, and no user has rated a significant fraction let alone all of the movies. The actual Netflix competition data is on the order of billions of entries and very sparse, and therefore presents a significant computational problem. Although the Netflix challenge has already

been conquered, we believe that some of the iterative methods used for collaborative filtering can be modified and parallelized to become significantly more efficient.

This paper is organized as follows. In section two we present our collaborative filtering model for users' movie preferences. The Alternating Least Squares (ALS) algorithm is presented in section three. Results from a small simulation study are presented in section four. Finally, some concluding remarks and opportunities for future work are given in section five.

THE MODEL

Let the r_{ui} denote the u th user's rating of the i th movie, for u in $\{1, \dots, U\}$ and i in $\{1, \dots, I\}$, where U and I are the numbers of users and movies, respectively. The r_{ui} are allowed to be any real number. A higher rating represents a more favorable score for a movie, as usual. Let \hat{r}_{ui} denote the predicted ratings. The collaborative filtering model from Bell and Koren's paper [1] is given by

$$\hat{r}_{ui} = \mu + a_i + b_u + p'_u q_i$$

EQN. 2.1

The overall objective of the model is to have the smallest mean squared error (MSE) between r_{ui} and \hat{r}_{ui} . The parameter a_i models the difference between a movie's rating and the average movie rating, for $i = 1, \dots, I$. Similarly, b_u represents the difference between a user's rating and the average user's rating for a particular movie for $u = 1, \dots, U$. The global average rating of all movies by all users is μ . Furthermore, the model takes into account the vectors of characteristics (violence, comedy, etc.) p_u and q_i , each of length (or dimension) d . Each element of p_u gives a rating of user's affinity for a certain characteristic, and the corresponding element in q_i quantifies the amount of this characteristic in the movie. The quantities, a_i , b_u , and μ are idealized modeling elements, which can be measured only if ratings by all users for every movie are known. These constitute the parameters of the model and need to be estimated using the available data.

Consider a simple case scenario where a recommender system is to determine whether a given user $u = 1$ will enjoy movie $i = 1$.

The user has rated several movies in the past generating a characteristic vector p_i , where $d = 3$. Suppose the vector measures comedy, action, and horror on a scale from -5 to 5 . If user u_i likes comedy, loves action, but hates horror, his characteristic vector may be in the form $p_i = (3.897, 4.875, -2.456)$. The characteristic vector for $i = 1$ is also of size $d = 3$ and is measured on comedy, action, and horror; suppose it is $q_i = (4.132, 3.978, -4.978)$. By taking the dot product of the two characteristic vectors, the recommender system can measure if the movie would be a good fit. This dot product varies from a minimum of -75 to a maximum of 75 . In our example, $p_i q_i = 47.61$ which means that movie $i = 1$ may be a good choice for user $u = 1$. Note that it is up to the recommender system to determine the d characteristics through the data, and that they may or may not have simple interpretations as in this example. Let

$$\kappa = \{(u, i) : \text{there is a rating } r_{ui} \text{ in the dataset}\}$$

for whichever dataset is being considered. To fit the parameters of the model (2.1) we define the objective function as

$$Q(\mu, a_i, b_u, p_u, q_i) := \sum_{(u,i) \in \kappa} (r_{ui} - \hat{r}_{ui})^2 + \lambda \left(\sum_{u=1}^U \|p_u\|^2 + \sum_{i=1}^I \|q_i\|^2 + \sum_{u=1}^U b_u^2 + \sum_{i=1}^I a_i^2 \right)$$

EQN. 2.2

The first term of the objective function represents the sum-squared error (SSE) between the true ratings and the fitted ones. The other term, with coefficient λ , penalizes parameters with large magnitude, and drives the algorithm toward a more parsimonious solution. Inside this penalty term is the sum of the Euclidean norms of the p_u and q_i vectors and the squared sums of the linear bias parameters, a_i and b_u .

FITTING THE MODEL

NOTE THAT THE problem presented in (2.1) and (2.2) is non-linear because of the presence of term $p'_u q_i$. Also, the objective function is not a simple sum of squared errors, because it includes a penalty

component. We use a method known as Alternating Least Squares (ALS) to minimize the objective function (2.2). As we shall see in the next section, ALS is effective in fitting this collaborative filtering model (2.1), and it is also amenable to simplifications to increase computational efficiency.

Note that the model is linear in a_i , μ , b_u , and p_u for a fixed q_i . Thus, for a fixed q_i the best linear unbiased estimate of a_i , μ , b_u and p_u can be determined by the Gauss-Markov Theorem (GMT) [2]. Also, the model is linear in a_i , μ , b_u and q_i for a fixed p_u . Similarly, for a fixed p_u the best linear unbiased estimate of a_i , μ , b_u and q_i can also be determined by the GMT. The ALS method is an iterative algorithm that alternates between solving for p_u and q_i and updates the parameters until MSE converges within a given tolerance or the iteration count reaches the maximum iterations. We begin the ALS algorithm by filling the characteristic vectors q_i with randomly generated real numbers ranging from -10 to 10. The quantities a_i , μ , b_u and y_{ui} are initialized by the following equations using the ratings given in the test data set:

$$\begin{aligned}\hat{\mu} &= \frac{1}{K} \sum_u \sum_i r_{ui} \\ \hat{a}_i &= \frac{1}{M_i} \sum_{u \in M_i} r_{ui} - \hat{\mu} \\ \hat{b}_u &= \frac{1}{N_u} \sum_{i \in N_u} r_{ui} - \hat{\mu} - a_i \\ y_{ui} &= r_{ui} - \hat{\mu} - \hat{a}_i - \hat{b}_u\end{aligned}$$

where K is total number of ratings, N_i is number of ratings given by user u and M_i is number of ratings given to item i . In the last equation, the variable y_{ui} is created by subtracting a_i , b_u , and μ from the ratings. This is done as an attempt to adjust these parameters out of the data in order to simplify our set of regression equations. In ALS, solving for p_u and q_i is done in two phases. In the first phase, we fix q_i and solve for p_u . In the second phase, we alternate and fix p_u and solve for q_i . The ALS algorithm continues to iterate between the two phases solving for p_u and q_i on the training data set until each parameter converges or the tolerance and/or iteration maximum is met. The pseudo-code corresponding to our implementation of ALS is on the following page.

Algorithm 1: ALS

```
Set up vectors  $y_{Users}$ ,  $y_{Items}$ ,  $x_q$ , and  $x_p$ 
Initialize  $\mu$ ,  $a_r$ ,  $b_u$ ,  $y_{ui}$ 
  while MSE has not converged and current iteration < max iteration limit
    Create  $q_i$  vectors
    Solve for  $p_u$  vectors
    Create  $p_u$  vectors
    Solve for  $q_i$  vectors
    Rescale each  $p_u$  and  $q_i$  vector
    Update  $\mu$ ,  $a_r$ ,  $b_u$ ,  $y_{ui}$ 
    Find MSE
End
```

We have not yet mentioned how λ is chosen. In practice, the optimal value of λ is obtained by the method of cross-validation. To implement cross-validation, the available data are divided into two disjoint sets. The training data, usually the larger of the two sets, is used for fitting the model and obtaining candidate sets of parameter estimates. The test data, a smaller subset but large enough to represent the sufficient diversity in the data, is then used to evaluate the candidate sets of parameters suggested by the training data. The measure of quality of fit used is usually the Mean Squared Error (MSE) or its square root, namely the Root Mean Squared Error (RMSE). In our simulation, the training and test datasets were about the same size, however some sparseness was imposed on the test dataset by removing one of the ratings per user.

Notice that the model simplifies when $d = 0$ and $\lambda = 0$. With these values, the p_u 's and q_i 's drop out, and the objective function (2.2) becomes a simple sum of squared errors. In this case the algorithm should converge in one step, although the fit will most likely not be adequate to make good recommendations.

SIMULATION AND RESULTS

IN OUR SIMULATION study, each data set contained four users and four movies, treating the two characteristics “comedy” and “violence” as relevant. We selected four movies at different ends of these spectra: *MacGruber*, *National Lampoon's Christmas Vacation*, *Citizen Kane*, and *Platoon*. For the purpose of this simulation we consider all four movies to be of the same overall quality, and let

$a_1 = a_2 = a_3 = a_4 = 0.01$. *MacGruber* and *Christmas Vacation* are comedies, so we set $p_{11} = p_{21} = 1$ and use $p_{31} = p_{41} = -1$ for the others. For the violence attribute, *MacGruber* and *Platoon* should rate higher and so we set $p_{12} = p_{42} = 1$, and use $p_{22} = p_{32} = -1$ for the others. We also consider four users whose preferences lie in the same areas of the comedy/violence spectrum, and whose rating biases are all the same (so that $b_1 = b_2 = b_3 = b_4 = 0.01$). We select $\mu = 2.5$ as the overall average rating. From this scenario, the rating data is generated using

$$r_{ui} = \mu + a_i + b_u + p'_{ui}q_i + \varepsilon_{ui}$$

for $i = 1, \dots, 4$ and where $\varepsilon_{ui} \sim \text{Normal}(0, 0.25^2)$. We selected several (u, i) combinations as missing, so that users have rated only two or three movies. The data was generated from supplied values of model parameters, according to Table 1.

TABLE 1: Parameters used in simulation study

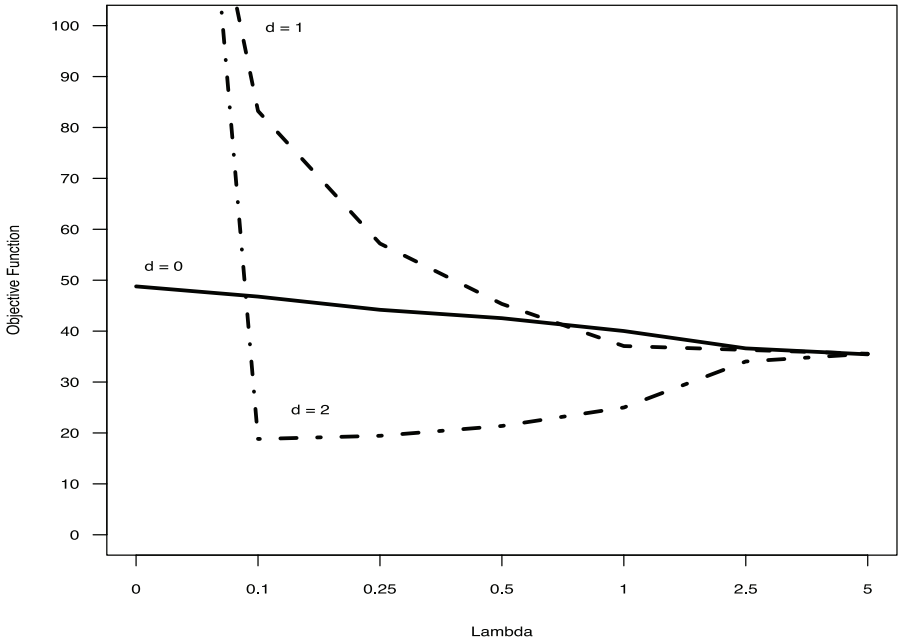
Movie	a_i	q_{i1}	q_{i2}	Movie Descriptions
1. <i>MacGruber</i>	0.01	1	1	violent comedy
2. <i>National Lampoon's Christmas Vacation</i>	0.01	1	-1	non-violent comedy
3. <i>Citizen Kane</i>	0.01	-1	-1	non-violent serious
4. <i>Platoon</i>	0.01	-1	1	violent serious
User	b_u	p_{u1}	p_{u2}	Prefers
User 1	0.01	1	1	violent comedy
User 2	0.01	1	-1	non-violent comedy
User 3	0.01	-1	-1	non-violent serious
User 4	0.01	-1	1	violent serious

We considered fitting the collaborative filtering model with ALS using three different dimension values for the characteristic vectors, $d = 0, 1, 2$ (where $d = 2$ is the true dimension). For each dimension, we tried seven settings of $\lambda = 0, 0.1, 0.25, 0.5, 1, 2.5, 5$. We refer to a particular (d, λ) combination as a "trial". Within each trial, 1000 randomly generated datasets are fitted with ALS; each fitting is referred to as a "subtrial". In each trial, the objective function, RMSE and MSE were calculated for each parameter over 1000 subtrials. We expect that over many subtrials, the estimates for a good model will converge to the true parameters.

TABLE 2: Simulation Mean of Q, RMSE and from 1000 subtrials

d	λ	Objective Function	Root MSE	μ
0	0.00	48.78433	1.730969	3.010583
	0.10	46.78605	1.691463	3.023322
	0.25	44.18945	1.641922	3.007001
	0.50	42.54068	1.606027	3.002809
	1.00	40.02031	1.553134	3.002197
	2.50	36.61012	1.483776	2.996164
	5.00	35.43677	1.463221	3.014009
1	0.00	10452.82	6.804822	1.607496
	0.10	83.23851	2.163152	2.877247
	0.25	57.20372	1.788586	2.908591
	0.50	45.33934	1.549798	2.935442
	1.00	37.06545	1.320608	2.971867
	2.50	36.53776	1.261688	2.985529
	5.00	35.55536	1.448333	3.023281
2	0.00	2180676	31.93812	0.538571
	0.10	18.81492	0.897336	3.010583
	0.25	19.44293	0.883624	3.008818
	0.50	21.37288	0.877922	2.979148
	1.00	24.99759	0.892402	2.999894
	2.50	34.03622	1.130104	3.010499
	5.00	35.54957	1.447850	3.024058

GRAPH 1: Plot of Q over changes in λ



When $d = 0$, as in the first set of trials, the average rating is based on a and b and does not contain user or movie characteristics. The lowest objective function result in the $d = 0$ trial set was 35.43. The second set of trials used p and q vectors of length $d = 1$. The objective function and RMSE also decrease with λ , but not uniformly. The change between $\lambda = 0.5$ and $\lambda = 1$ is more apparent than $\lambda = 2.5$ and $\lambda = 5$. The final seven trials were run at the true dimension of p and q , $d = 2$. These trials also reported a better objective function and RMSE for similar λ values than the other trial sets. The only case in which this is not true is $\lambda = 0$, where none of the results are very good. Note in this case, the objective function decreases initially and increases for larger values of λ . The bowl shape of the objective function is typically encountered during cross-validation. Trial number two ($\lambda = 0.1$), seemingly has the best balance between objective function and μ . For each trial set, $\lambda = 0$ always reported the worst objective function. This is because of the nature of our objective function. Referring back to equation (2.2) shows how $\lambda = 0$ cancels out a significant portion of the objective function. Thus, λ allows for a trade-off between the sum of squared error and the magnitude of the parameters. That is, sacrificing a small amount of prediction accuracy allows us to obtain

a lower objective function. Something else to observe is the final trial for each set. Comparing the trend in the attained minimal values of the objective function for the various values of λ considered in the study, we note that for $d = 2$, the best overall value occurs for $\lambda = 0.25$. However, for $d = 0$ and $d = 1$, the value of attained minimum of objective function seems to be decreasing as λ increases. Note that the true model underlying our simulations has $d = 2$. Clearly, misidentifying the dimensionality (d) of the characteristic vector of the collaborative filtering model does impact the estimation process significantly.

CONCLUSIONS

WE CONSIDERED THE estimation of a collaborative filtering model using an algorithm known as alternating least squares (ALS). The purpose of the article is to provide an overview of the collaborative filtering model and give the flavor of the estimation methodology. From the results of the small set of trials presented here, we conclude that the ALS algorithm accurately estimates the parameters in our model. Using different dimensions of d also confirmed that knowing the true dimension is important in fitting the model. However the scope of our simulation in this study is very limited. The next step should be to expand the scope of the simulation to include larger data sets and a variety of parameter set ups. More work is also needed to address the issue of determining the dimension (d) of the characteristic vectors.

ACKNOWLEDGMENTS

THE HARDWARE USED in the computational studies is part of the UMBC High Performance Computing Facility (HPCF). The facility is supported by the U.S. National Science Foundation through the MRI program (grant no. CNS--0821258) and the SCREMS program (grant no. DMS--0821311), with additional substantial support from the University of Maryland, Baltimore County (UMBC). See www.umbc.edu/hpcf for more information on HPCF and the projects using its resources.

The work is an extension of the author's project during the High Performance Computing Research Experience for Undergraduates (REU) program (www.umbc.edu/hpcreu). The author gratefully acknowledges his teammates Julia Baum (Worcester Polytechnic Institute), Cynthia Cook (Catawba College), Joshua Edgerton (Cornell University), and Scott Rabidoux (Wake Forest University). We gratefully acknowledge the funding and the mentoring we received from our graduate student mentor Mr. Andrew Raim, off-campus client Dr. Robert Bell, AT&T Bell Labs, and faculty mentor Dr. Nagaraj Neerchal.

WORKS CITED

1. Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.
2. Raymond H. Myers. *Classical and Modern Regression with Applications*. Duxbury Press, 2nd edition, 2000.

FORCE IN STASIS

Spectral, Spatial, and Theoretical Findings in “Pilentze Pee”

samuel garrett

Samuel Garrett is a composer and guitarist based in New York City. His music is characterized by a focus on the aesthetic relationship between rhythmic simplicity and complexity, dynamic melodic and harmonic stasis, and unique timbral composites. Garrett studied music composition with Drs. Linda Dusman and Stuart Saunders Smith at University of Maryland, Baltimore County, where he graduated summa cum laude and with departmental honors in December 2011. He is the recipient of multiple New Music Ensemble fellowships and the Johann E. Eltermann Scholarship for 2010-2011. Garrett looks forward to establishing an ensemble dedicated to the performance of his work and to pursuing graduate studies in music composition. Garrett thanks Dr. Linda Dusman for her mentorship and insight during his studies.

I was drawn to the recording of “Pilentze Pee” from *Le Mystère des Voix Bulgares* because of its idiosyncrasies. This paper was my opportunity to carry out an investigation of idiosyncratic musical character – to find and identify aspects that distinguish this piece and render it aesthetically compelling. My analysis of “Pilentze Pee” ultimately led to the unification of two disparate interests – research into spectral analysis and musical proportionality and an interest in harmonic stasis. As a composer, I am attentive to the interaction of these properties and therefore welcomed the opportunity to explore their qualitative functions in “Pilentze Pee.” My analysis constitutes primarily spectral and proportional studies that illuminated several fascinating aspects of “Pilentze Pee.” This paper is my first foray into academic research, and I am proud to present it to the *UMBC Review* audience.

INTRODUCTION

THE MUSIC OF *Le Mystère des Voix Bulgares* (*The Mystery of the Bulgarian Voices*), recorded by the Bulgarian State Radio and Television Female Vocal Choir, is emblematic of the methodological and theoretical discontinuities between music theorists and ethnomusicologists (Webster 168). From one perspective, the music evinces a rich and attractive musical character, as reflected by its successful commercial release on record labels 4AD and Nonesuch. From another perspective, the music represents a gentrified, indistinct, and commoditized national product (Silverman 212; Rice, *May It Fill* 28). While it remains important to note the intrinsic antagonisms between this music and world culture, this paper seeks to avoid such conjectural debates. Rather, it discusses the mystery of this music largely outside of the Western musical theoretical lens. Using “Pilentze Pee” from *Le Mystère des Voix Bulgares* as representative of the Bulgarian aesthetic, I conducted a series of spectral, spatial, and other musical analyses to illuminate the varied or nuanced qualitative responses to Bulgarian vocal music in general and the six-part a cappella arrangements in *Le Mystère des Voix Bulgares* recordings specifically.

Daniel Harrison, in *Harmonic Function in Chromatic Music: A Renewed Dualist Theory and an Account of Its Precedents*, writes on harmonic stasis: “In the absence of other harmonic stimuli, the only [harmonic stimulus] present tends to arrogate Tonic unto itself... Thus, some harmonic entity that seems immobile naturally attracts consideration as Tonic and actually can attain that status if other counterbalancing factors...are not at work or are overwhelmed” (Harrison 80-81). By Harrison’s definition, the drone-heavy and modal “Pilentze Pee” exists suspended in harmonic stasis, framed and reinforced by a persistent and repetitive rhythmic scheme. Similar melodic and harmonic content reiterates itself with only slight variation throughout the strophic musical form, as later exemplified in examples 1 and 2. This compositional stasis raises two questions: why does this music radiate as such in a static musical environment and what are the other counterbalancing factors that cause “Pilentze Pee” to remain so powerfully evocative? By looking deeper into the fabric of “Pilentze Pee” and past the Western theoretical analytic construction of stasis, this paper reveals various elements that contribute to the unique experience of this music.

FORCE IN STASIS

AFTER AN INITIAL listening of “Pilentze Pee,” I arrived at the concept of force in stasis. Force in stasis refers to the elements I have found in this piece that render it qualitatively and aesthetically charged — a quality felt through perceived and unrelenting elements of musical tension. These variables, acting in oppositional force against the static form, create a musical equilibrium that distinguishes “Pilentze Pee.” Notable spectral idiosyncrasies, temporal and spatial proportional relationships, and melodically and harmonically charged pitch content constitute the force in stasis characteristic of “Pilentze Pee.”

GLOBAL MUSICAL FORM AND TEMPORAL PROPORTIONS

“PILENTZE PEE” CONSISTS of seven sections (I-VII), six of which are repetitions with slight variations of a core diaphonic melody (see ex. 1) that occurs during mm. 1-6 (0.0000 to ~12.7245 seconds). The piece ends with a short cadential section (VII), a measure of 9/4. These sections are punctuated with moments of silence, adhering to scored fermatas and rests, which endure the amount of reverb decay applied on the recording.



EX.1 Core Diaphonic Melody; Subsections *A* (top image) and *B* (bottom image).
Source: Krassimir Kiurkchiiski, “Pilentse Pee: A Nightingale Sings,” (Voxbulgarica, 2011). Transposition by author.

Each of these six sections contains two subsections, *A* and *B*, also punctuated with moments of silence. Subsection *A* consists of three 4/4 measures, and subsection *B* consists of three 9/16 measures (see ex. 1). The last cadential section, VII, utilizes only subsection *A* — differing from the rest of the global rhythmic scheme — in a time signature of 9/4 (see table 1).

Sections	Subsection A	Subsection B
I-VI	3 measures of 4/4	3 measures of 9/16
VII	1 measure of 9/4	

TABLE 1 Rhythmic and Structural Constituents of Sections I-VII.

The final structural result consists of sections I-VI, each constituted by two internal subsections, followed by section VII. These sections and subsections are visually represented in “Pilentze Pee” as a simple waveform (see fig. 1), and their composite (subsections *A* and *B*) is outlined in fig. 2.

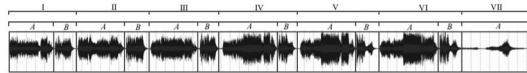


FIG. 1 Global Waveform.

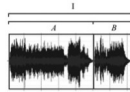


FIG. 2 Waveform of mm. 1-6 (0.0000 to ~12.7245 seconds); two-part diaphonic section.

Rhythmically, a charged sort of tension and release characterizes this piece. Though all variations of the strophic form of “Pilentze Pee” occur within the melody and harmony, the rhythmic material remains identical. How can such tension be created in a repetitive structural form?

The variation in pulse from a quarter note to a 16th note between subsections *A* and *B* in even to odd meters respectively creates an asymmetrical aesthetic, imbuing an enduring sense of tension in the otherwise static musical scape of “Pilentze Pee.” Using a common rhythmic pulse of a 16th note, detailed research into the proportional relationship between these distinct and repeating subsections reveals an interesting local and global proportional scheme. The following analysis utilizes a 16th note pulsation as the common element of duration since the metric feel of subsections *A* and *B* does not necessarily reveal their respective meters to the listener. The rhythmic content from mm. 1-6 and the proportional metric rhythmic relationship scheme between subsections *A* and *B* for the entire piece, aside from section VII, appear below (see table 2).

<i>Number of Measures</i>	<i>Rhythmic Attributes of Subsection A</i>	<i>Rhythmic Attributes of Subsection B</i>
3 [each]	4/4; 4 sets of ♩ per one measure; total of 16 ♩ per measure; 16x3=48 ♩ per rhythmic scheme	9/16; 9 ♩ per one measure; 9x3=27 ♩ per rhythmic scheme

TABLE 2 Rhythmic Content as 16th Note Pulsations and Resulting Proportional Schemes.

This research reveals a rhythmic proportional scheme of 27:48 (*B:A*), barring some exceptions in variation. Using the proportional relationship of 27:48 (0.5625) as almost the direct halfway point between two common and psychoacoustically attractive proportions — .5 or 50% and the Golden Ratio of .618 — presents an intrinsic asymmetry, at least to the Western-trained ear. Using the scheme of subsection *A* followed by subsection *B* (longer followed by shorter rhythmic durations) throughout the piece synthesizes a high sense of rhythmic and proportional tension alone. This scheme, coupled with a proportionality centered between two common musical proportions, creates a stretched tension, enriching the simple, or static, global organizational form.

Further research brings to the fore a yet more complex global scheme in “Pilentze Pee.” The total sum of sections I-VI paired with the sum of the additional 16th note pulse of section VII (see table 3) unveils a remarkable proportional resolution — a proportional release to the persistent asymmetrical tension throughout sections I-VI.

<i>Sections</i>	<i>Rhythmic Attributes of Subsection A</i>	<i>Rhythmic Attributes of Subsection B</i>
I	48 ♩	27 ♩
II	48 ♩	27 ♩
III	48 ♩	27 ♩
IV	48 ♩	27 ♩
V	48 ♩	27 ♩
VI	48 ♩	27 ♩
VII [cadential section; 9/4]	9/4; 9x4=36 ♩	0 ♩

TABLE 3 Rhythmic Attributes as 16th Note Pulsations in Sections I-VII.

The total sum of all rhythmic attributes of subsection *A* (288) and the total sum of all rhythmic attributes of subsection *B* (162) are proportionally related as described above (.5625). Adding section VII to the sum of 16th note pulses in subsection *A* (288+36=324) redefines the proportional relationship as 162:324, showing an interesting global 16th note pulse value (see table 4).

Sections	Subsection A	Subsection B	Proportional Relationship (B:A)
I-VI	48 [♯]	27 [♯]	162:288=.5625
+VII	36 (288+36) [♯]	0 [♯]	162:324=.5

TABLE 4 Proportional relationships of subsection A and B (B:A) in Sections I-VI and Section VII.

The importance of section VII in this redefinition of global pulse proportion as a cadence cannot be overstated. With the inclusion of the pulse values of section VII, the final rhythmic proportions of the piece shift from an insistent asymmetry, awkwardly sitting between .5 and .618, to the well-established proportion of .5 — a proportional tension and release.

The global form of “Pilentze Pee” therefore remains a static and comfortable strophic structure. Yet, the juxtaposition of a micro and macro-asymmetry and a wholly symmetrical conclusion is difficult to ignore. This raises two questions: why would the asymmetry of sections I-VI be so precisely and awkwardly designed between .5 and .618, and, moreover, why did the composer decide to conclude the duration with the additional measure of 9/4 (section VII) and subsequently reach the proportional unity of .5?

The stark contrast between global form and temporal proportions of “Pilentze Pee” reveals an awareness of tension and release in a static environment. The forces of a simple global form and a seemingly strategically complex asymmetrical tension followed by its inevitable release act in total equilibrium, charging musical stasis with rich and complex proportional undertones. The psychoacoustic effect of this proportional device demands further investigative analysis.

AN INTRODUCTION TO SPECTRAL ANALYSIS

ROBERT COGAN, IN *New Images of Musical Sound*, writes, “parallel to the universe that we see and touch exists a sonic universe that we hear. Beneath the surface of that sonic world, and in music beneath the surface of notes and instruments, there exists another world: that of sound waves, of overtones or partials that make up the sound spectrum, and of other minute, obscure phenomena” (4). The spectrogram attempts to address the force with which “Pilentze Pee” is charged beyond Western melodic and harmonic theoretical analy-

sis. The analytical usefulness of the spectrogram into the uncharted is unparalleled. Cogan states, “[spectrograms] objectify much that has been most elusive, even mystifying, about sounds and the ways they create the design of musical structures. In so doing, they illuminate the very nature of musical structure and expression” (3). Indeed, there remains no better way to analyze and address *Le Mystère des Voix Bulgares* than through a method known for its window into the unknowable qualities of music.

SPECTRAL ANALYSIS OF “PILENTZE PEE”

USING THE PROGRAM Sonic Visualiser, I produced a series of spectrograms of the popular Bulgarian State Radio and Television Female Vocal Choir recording of “Pilentze Pee” from *Le Mystère des Voix Bulgares* (see figs. 3 and 4).

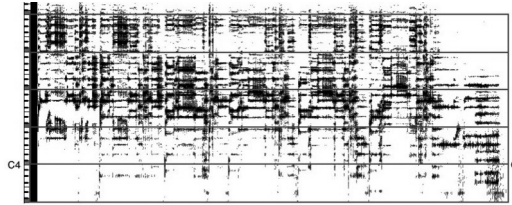


FIG. 3 Spectral classes in “Pilentze Pee.”

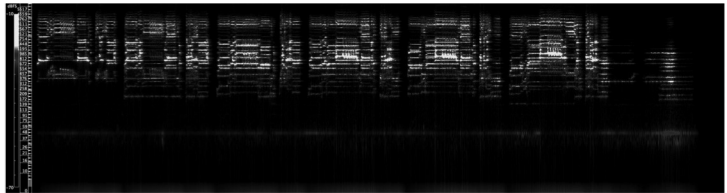


FIG. 4 Spectrogram of “Pilentze Pee” in totality.

The spectrograms, at first glance, appeared incorrect due to the unusually high amplitude of the higher partials and overtones. After adjusting the spectral parameters, the spectral acoustical idiosyncrasies became apparent. The spectral content resonates at the first overtone or second harmonic, an octave above the fundamental (see figs. 3 and 4 for details of frequency ranges and overtones). This

spectral power prevails as the dominant characteristic and remains present throughout. Why is this the dominant force throughout “Pilentze Pee,” and how does it charge the relatively static sonic environment?

The answer begins with an aspect as simple as amplitude: loudness functions as the most characteristic and present dynamic quality in “Pilentze Pee.” The loud dynamic rarely ceases throughout “Pilentze Pee” and nearly saturates the entire frequency spectrum, the highest amplitude residing in the first overtone and higher partials. The loud dynamic overwhelms the spectrum of higher partials so much so that the fundamental pitch content appears to be a mere reflection of the richness of the higher partials.

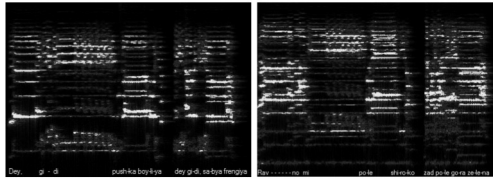


FIG. 5 Spectrogram of Section I (left figure) and spectrogram of Section II (right figure).

The two spectrograms above (see fig. 5) reflect the spectral content of sections I and II. Section I consists of two-part diaphonic singing; section II employs the entire six-part ensemble. At first glance, a greater number of partials between section I and II is clear, though the difference between the two is not particularly overwhelming. The music remains as spectrally intense with just two active parts as it does with six. Volume allows for a general push into higher frequency spectra, permitting an incredible spectral sheen to coat the music. Of particular interest is the glissandi section, occurring on the syllable “di [‘ee’].” In the spectrogram, note the spiraling of frequency content at the lower bands as well as an activation of the higher partials. However, volume is not the sole contributor to the spectral idiosyncrasies between the spectrograms in fig. 5.

The vowel sounds in the lyrics of “Pilentze Pee,” especially in “di” (as in “see”), “mi” (as in “me”), and “pole” (as in “lay”) are incredibly powerful, unveiling the highest represented partials in the spectrogram. The article “Resonance strategies used in Bulgarian women’s singing style: a pilot study” begins to address the question of whether these vowel sounds and their resulting spectra are so intensely charged due to volume alone. The appearance of higher spectra —

most notably the first overtone — in “Pilentze Pee” likely emanates also from a resonance tuning strategy used in the Bulgarian style (Henrich et al. 3-4). In “Resonance strategies,” researchers looked at the frequency content and measurements of vocal tract resonances of the two characteristic singing styles of Bulgarian women — *tesbka* (loud, heavy) and *leka* (light, quiet) — in order to address the “unexpected strength of the second harmonic and the overall loudness” (Henrich et al. 3). The researchers concluded that “Bulgarian singing technique usually tunes the first resonance (R1) to the second harmonic (H2) for most vowels” and that “[t]uning of the vocal tract resonance contributes to the predominance of the second harmonic in Bulgarian women’s singing in these styles, to its characteristic timbre and to the high sound levels achieved” (Henrich et al. 7).

The research above provides integral detail of the force in “Pilentze Pee”: an idiosyncratic timbre, constituted by not only volume but also a resonant tuning strategy.

WESTERN MUSICAL ASSOCIATIONS

I HAVE FORGONE a formal analysis of “Pilentze Pee” in order to focus primarily on notable spectral and spatial characteristics. That stated, a brief application of a Western theoretical lens further underscores the idiosyncratic characteristics of “Pilentze Pee.”

In typical Bulgarian vocal and instrumental music, a modal two-part texture is employed in which one vocalist or instrument performs a melody against a drone-heavy accompaniment (Rice, *Polyphony* 16-19). The closely-voiced sonorities are dissonant due to the concentration of intervals of a second, a fourth, a fifth, a seventh, and a ninth. These intervals are uncommon in common practice period Western vocal music, which favors the third and the sixth as defining melodic and harmonic qualities. The intrinsic tonal language of this music colors its stasis, informed by its general orbit around a central modal tonic.

Another interesting aspect of Bulgarian vocal music resides in performance: this music does not utilize the pulsating pitch effect of vibrato. Besides gestures, which employ glissandi, the vocal sheen of *Le Mystère des Voix Bulgares* radiates through clarity and precision. Whereas vibrato often serves as a technique to evoke emotion in performance, volume and the distinct Bulgarian timbre are largely the only affected devices used in this music. The lack of vibrato helps

to intensify the closeness of the voicings, effectively highlighting the idiosyncratically dissonant tonal vocabulary.

In arrangements such as “Pilentze Pee,” geographic variances in Bulgarian folk music are assimilated into a homogenous national product (Rice, *May It Fill* 28). Instead of having the typical two-part texture, these arrangements are written for six-parts and sung a cappella by a choir, such as the Bulgarian State Radio and Television Female Vocal Choir.

SPACE AND SILENCE

THE PSYCHOACOUSTIC EFFECT of silence has long been used as a device for building intensity in music, especially in music that is traditionally heard as static (Metzer 332). Within the architecture of “Pilentze Pee” exist a number of effective devices of space and silence used in the static sonic environment. The device explored here revolves around the relationship of repetitive silences, notated as fermatas and rests, and a distinct rhythmic break employed in the final section, VII.

As seen in ex. 1, each section employs three fermatas in its subsections: the first of these fermatas occurs in the second measure of each subsection *A*, acting as a sub-temporal device which stretches a held pitch. The second occurs at the end of subsection *A*, dividing the two subsections. The third occurs at the penultimate measure of subsection *B*, stretching the measure before a brief three-note cadence. A final silence consisting of seven 16th note rests (using the established 16th note pulse previously noted) concludes subsection *B*. Combining the final silence with the preceding three-note cadence, these temporal stretches and silences, when fragmented into the 16th note pulse utilized in the analysis above, present an interesting phrase scheme that occurs throughout sections I-VI and fractures and redefines itself during section VII. This phrase scheme, pulsed in 16th notes, is a set of 28, 20, 17, 10 (2+7) (see table 5), each section employing a gradual accelerando in phrasing, silences, and stretches.

Sections	Phrase Scheme in 16 th Note Pulsations
I-VI	28 [♯] / 20 [♯] / 17 [♯] / 10 [♯]

TABLE 5 Phrase Scheme in Sections I-VI.

Having identified the silence system and phrase scheme within each section, it is important to note the pervasive repetition, ensuring

that the listener will not forget the pattern of silences and stretches throughout the piece.

Section VII, the final cadential section, employs a fascinating and effective device: a fragmentation of the phrasing pattern outlined above. This fragmentation — a fermata and brief pause — occurs four quarter notes into the final 9/4 measure that, when translated into the 16th note pulse, equals sixteen 16th notes. This follows the completion of a normal subsection *A*, re-metered in 9/4 and utilizing the final five quarter notes. When the second half of section VII is divided based on the entrance of voices 3, 4, 5, and 6, a unique and surprising symmetry appears (see ex. 2 and table 6).



EX. 2 Section VII. Source: Krassimir Kiurkchiiski, “Pilentse Pee: A Nightingale Sings,” (Voxbulgarica, 2011). Transposition by author.

Voices	9/4; Division I: Beats 1-4	9/4; Division II: Beat 5	9/4; Division III Beats 6-9
1	4x4=16 [♭]	1x4=4 [♭]	4x4=16 [♭]
2	4x4=16 [♭]	1x4=4 [♭]	4x4=16 [♭]
3	Rest	Rest	Whole-note entrance; 4x4=16 [♭]
4	Rest	Rest	Whole-note entrance; 4x4=16 [♭]
5	Rest	Rest	Whole-note entrance; 4x4=16 [♭]
6	Rest	Rest	Whole-note entrance; 4x4=16 [♭]

TABLE 6 Entrances and Resulting Phrase Scheme in Section VII.

Again, a strong symmetry appears out of an inherently asymmetrical pattern: a silence system of 28, 20, 17, and 10 pulses per section is suddenly fragmented, creating a silence and phrasing pattern of 16, 4, and 16 with which the piece concludes. The numbers 16, 4, and 16 represent a moment of symmetrical mathematical purity in an asymmetrical work — the square and the square root (see table 7).

Sections	I-VI	VII
Phrase Scheme	28 [♯] / 20 [♯] / 17 [♯] / 10 [♯]	16 [♯] / 4 [♯] / 16 [♯]

TABLE 7 Phrase Scheme Comparison Between Sections I-VI and Section VII.

SYNTHESIS OF FORCE IN STASIS

A REVIEW OF the elements at work in “Pilentze Pee” reveals that the internal and opaque qualities of this piece render it anything but static. The proportional asymmetry and ultimate symmetrical resolution unveil a hidden force driving this music toward its ultimate proportional cadence. The overwhelming and dominant presence of the first overtone and higher partials found in spectral analysis reveals not only empirical data on the frequency content of vocal loudness (tuning systems), but also how these aspects color and constitute the unique spectral timbre associated with this music (Henrich et al. 3-4). These veiled qualities synthesize a powerful and complex intensity, perhaps the very color and intensity which makes *Le Mystère des Voix Bulgares* so attractive. In the above discussion on Western theoretical elements, the pervasive use of seconds, fourths, fifths, sevenths, and ninths creates a tense and haunting set of sonorities.

All of these elements constitute a uniquely charged static and modal music — a force in stasis. This research unveils layers of complexity hidden in “Pilentze Pee,” providing a reference for what lies beyond the Western theoretical analysis of this music as static.

WORKS CITED

- Cogan, Robert. *New Images of Musical Sound*. Cambridge and London: Harvard University Press, 1984. Print.
- Harrison, Daniel. *Harmonic Function in Chromatic Music: A Renewed Dualist Theory and an Account of Its Precedents*. London: University of Chicago Press, 1994. Print.
- Henrich, Nathalie, et al. "Resonance strategies used in Bulgarian women's singing style: A pilot study." *Logopedics Phoniatrics Vocology* 32.4 (December 2007): 171-177. *Academic Search Premier, EBSCOhost*. Web. 19 Apr. 2011.
- Le Mystère des Voix Bulgares*. Perf. The Bulgarian State Radio and Television Vocal Choir. Nonesuch, 1990. CD.
- Metzer, David. "Modern Silence." *The Journal of Musicology* 23.3 (2006): 331-374. *JSTOR*. Web. 29 April 2011.
- Rice, Timothy. *May It Fill Your Soul: Experiencing Bulgarian Music*. Chicago: University of Chicago Press, 1994. Print.
- Rice, Timothy. *Polyphony in Bulgarian Folk Music*. Diss. University of Washington, 1977. Ann Arbor: UMI, 1977. Print.
- Silverman, Carol. "'Move Over Madonna': Gender, Representation, and The 'Mystery' of Bulgarian Voices." *Over the Wall/After the Fall: Post-Communist Cultures Through an East-West Gaze*. Eds. Sibelan Forrester, Magdalena J. Zaborowska, Elena Gapova. Bloomington, IN: Indiana University Press, 2004. 212-238. Print.
- Webster, Jamie Lynn. "The Mysterious Voice! American Women Singing Bulgarian Songs." *The Anthropology of East Europe Review* 22.1 (2004): 155-169. Print.

CROSSING BOUNDARIES

Foreign Language Learning for Students with Learning Disabilities

danielle viens-payne

Danielle Viens-Payne conducted this research as a modern languages, linguistics, and intercultural communication (MLLI) major during 2009-2010. She was accepted into the MLLI Honors Program in 2009 and initially worked with Dr. Thomas Field, who led her to another advisor, Dr. Ana Oskoz. Danielle worked with Dr. Oskoz to complete her thesis, which she then successfully defended in Spring 2010. She graduated from UMBC in December 2010 Magna Cum Laude with a B.A. in MLLI, departmental honors, and a Certificate of General Honors from the Honors College. Danielle would like to thank Dr. Field for his guidance and interest in her work and for leading her to Dr. Oskoz. Danielle would also like to thank Dr. Oskoz for her relentless dedication, expert guidance, and the countless hours sacrificed to help her learn and grow as a researcher.

My motivation for this research began when I took an education course that required the observation of a high school foreign language class. I noticed that any student with below average performance was left behind. This was never the intention of the teacher since she was not provided with the proper resources to teach these students, particularly students with learning disabilities. Ultimately, the school did not provide these students with adequate help to succeed in foreign language classes. Student support staff at UMBC reminded me that similar issues can exist at the university level. My research is therefore focused on uncovering potential ways to help students with learning disabilities succeed in foreign language classes in secondary schools and institutions of higher education. I shared this research with instructors and learning specialists, presented it to foreign language professors and student support staff at UMBC's URCAD in 2010, and am now pleased to share it with the *UMBC Review* audience.

INTRODUCTION

LEARNING A FOREIGN language (FL) is a graduation requirement in U.S. secondary schools and a number of U.S. institutions of higher education. This means that every year thousands of students fill up language classes to complete this requirement. Some of these students struggle with not only the difficult task of learning a new language but also a hindering factor that stands in the way of second language acquisition: a learning disability. This research focuses on FL learning within the state of Maryland for students with learning disabilities (LD). These students are defined in this paper as learners who have phonological and/or orthographic processing problems, such as dyslexia and dysgraphia. More specifically, this research focuses on FL requirements and policies for students with LD in Maryland secondary schools and institutions of higher education and on the potential programs and/or policies that could be implemented to assist students with LD in successfully learning a FL.

When examining the issue of FL learning for students with LD, one must take into account three problematic areas. First, learning disabilities range in type and severity and no two students with LD have the exact same issues in a particular classroom environment. Some students with LD have mild symptoms which may allow them to take FL classes with few or no problems with the coursework, while other students with LD struggle tremendously due to a more severe learning disability. A second problem, due to the varying types and severity of learning disabilities, is the difficulty of implementing a general rule or protocol for students with LD in the area of FL requirements. Many schools implement fairly universal accommodations for students with LD to which the instructors must adhere. Finally, the schools' accommodations are not always enough to help students with LD achieve success in the FL classroom, nor are they always enough to help the instructor teach students with LD. Balancing the importance of FL learning with the importance of helping students with LD to succeed is necessary.

In this interconnected world, where people from different cultures interact with one another on a regular basis, it is helpful to understand other cultures and to speak other languages. The educational system, which prepares students to excel in the real world, should provide all students, including those with learning disabilities, with the opportunity to be successful at learning a FL. This study reviews current policies and requirements for these students, explores

which policies are working and which could be improved, and suggests alternative policy structures. The research questions explored in this study are the following: What do Maryland FL instructors and learning specialists do to help students with LD? What support do Maryland schools provide to FL instructors and learning specialists to work with students with LD in FL classes? What do Maryland FL instructors and learning specialists believe should be done at the state level to support schools?

BACKGROUND

IN 1990, THE Americans with Disabilities Act (ADA) passed, mandating equal access and opportunities for Americans with disabilities and an end to discrimination based on disabilities (“ADA Home Page,” n.d.). This act is continually amended to effectively aid and protect individuals with disabilities (“ADA Home Page,” n.d.). In 2004, George W. Bush reauthorized the Individuals with Disabilities Education Improvement Act (IDEA), an act that specifically concerns the education of students with disabilities. IDEA defines the term *learning disabilities* as disorders “in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which disorder may manifest itself in the imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculations” (“IDEA 2004,” n.d.). According to IDEA, all students with disabilities, including those diagnosed with a learning disability, must be allowed access to a free public education and must have an Individualized Education Plan (IEP) developed for them according to federal guidelines.

A student’s IEP will identify his or her unique disability and describe the accommodations that must be made for the student. There are three general categories of learning disabilities: dyslexia (difficulty with reading), dysgraphia (difficulty with writing), and dyscalculia (difficulty with math). While these three categories seem to narrow down the various types of learning disabilities, they are actually much more complex because “[d]ifficulties vary in severity between individuals” (Crombie & Schneider, 2003, p. ix-x), confirming why it is difficult to implement one standard for all students. Thus, students with LD should be evaluated individually to diagnose the specific problem and its severity, as well as their scholastic performances in foreign language classes (“Learning Disabilities: An

Overview,” n.d.). Maryland public secondary school systems must adhere to IDEA, and institutions of higher education must adhere to ADA. Therefore, in K-12 institutions, students with LD must be evaluated and a plan of action must be formulated to aid the student by using state funding and school resources. In college, students must come forward themselves and request services. Once it has been proven and recorded that the student has a learning disability, the university must provide accommodations for the student.

A generic approach to assist in the teaching of FL to students with LD is the Universal Design for Learning (UDL). The Center for Applied Special Technology (CAST) defines UDL as the following:

Universal Design for Learning is a set of principles for curriculum development that give all individuals equal opportunities to learn. UDL provides a blueprint for creating instructional goals, methods, materials, and assessments that work for everyone – not a single, one-size-fits-all solution but rather flexible approaches that can be customized and adjusted for individual needs. (“CAST: About UDL,” n.d.)

While generic, UDL can help FL teachers design a few alternative methods of teaching foreign languages so as to address all learners in the classroom and not solely the average learner.

Policies specifically concerning FL requirements for students with LD have not been explored and addressed in depth. According to the policies posted on the websites of Maryland public schools, IDEA is followed by providing student support services programs, employing learning specialists to work with students with LD, and making accommodations for these students. The generic protocol is that students with LD may be exempted from completing required FL credits if the students are interviewed, their cases are reviewed, and it is found that the students have attempted a language course (and failed or done poorly) and show at least average performance in other courses (Spratt, personal communication, September 2009). An unwritten rule, though one of common knowledge among educators and learning specialists, states that students with LD must prove to be unsuccessful in the FL classroom before more individualized plans are put into place for them (Spratt, personal communication, September 2009). Some educational institutions require a replacement of the exempted credits, which means that

the exempted student must complete a different course to substitute for the FL course. FL instructors, learning specialists, and researchers agree that there should be a wider array of accommodations for students with LD in the area of FL learning. However, due to the overall lack of understanding of learning disabilities and the strategies needed to teach FLs to students with LD, policies in this area are still very broad.

RELATED STUDIES

ALTHOUGH THE NEEDS of students with LD have not received much attention, the 2009 issue of *Foreign Language Annals* illustrates the current interest in this area by addressing the connection between FL learning and learning disabilities (Sparks, 2009a; Sparks, 2009b; Leons, Herbert, & Gobbo, 2009; Amend, Whitney, Messuri, & Furukawa, 2009). This issue focused on problems that arise in the classroom, associated policies and requirements, and suggestions that could ensure that “the bigger picture [of] creating foreign language learners who experience success and want to continue the foreign language learning process” is still in sight (Young, 2009, p. 6).

On discussing problems, solutions, and challenges involved in FL learning for students with LD, Sparks (2009a) states that “in some cases, the language learning problems of students classified as LD are thought to be so severe that educators recommend that students receive waivers and/or course substitutions for the FL requirement,” and that there is “easy access to course waivers and substitutions” (p. 3-4). However, Sparks (2009a) continues, special education advocates promote the importance of the inclusion of special needs learners “in regular classes, including foreign language classes” (p. 4). Even though these advocates believe that it is important for students with LD to be included in a regular classroom environment, it is also important not to lose sight of the fact that these students are not average learners; they need more assistance and more resources at their disposal (Spratt, personal communication, September 2009). For instance, one method which helps students with LD learn vocabulary is Mastropieri and Scruggs’ keyword mnemonic method. This method associates two words with the new vocabulary term being taught; one word sounds like the new term being taught to the student, and the second word is the actual meaning of the new term. Therefore, when the student comes across the new vocabulary

term, the student remembers the first word that sounds like the new term, and then remembers the meaning associated with it. Their study shows that this method has positive outcomes among students with LD, helping them to learn and retain foreign language vocabulary words (Mastropieri & Scruggs, 1990). This method could be helpful to instructors in addressing one part of the foreign language learning process.

To address the needs of students with LD, Amend, Whitney, Messuri, & Furukawa (2009) conducted a two-semester long study in which they delivered a modified Spanish course. In this course, they reduced the amount of information taught to students with LD, focused on repetition, and emphasized tactile activities and metacognitive strategies. Teachers administered semester-end written essay exams in this modified course, as per the university's standard requirement. These essay exams were compared to essays written by students in non-modified courses to see if there were differences and if the modified course had helped the students with LD succeed any more than they would in a standard course. Two professors graded the anonymous exams from both the modified and unmodified courses. There were no differences between essay exams written by students with LD and essay exams written by average students. The researchers suggested that the tactics used in this modified course (including reduction of information taught, repetition, tactile activities, and metacognitive strategies) could help all students succeed in FL classes (Amend et al., 2009).

Leons, Herbert, & Gobbo (2009) researched teaching practices in the area of FL learning for students with LD, stressing that FL professors are not learning disability specialists, and that not all FL professors are familiar enough with learning disabilities to know how to help students with LD. According to their research findings, recommended practices for instructors teaching students with LD in FL classes include: being conscious of the curriculum and the pace of the learning, using multimodal teaching methods, crafting classroom activities to ensure student success, being supportive and encouraging, setting up one-on-one instruction or tutoring sessions, providing a structured class environment, viewing each student as unique with individual strengths and weaknesses, starting instruction at each student's point of readiness, and offering varying means of assessment (Leons et al., 2009, p. 50-51). This study states that accommodations made for students with LD "often miss the mark"

and encourages the use of these recommended practices to help both students with LD and their FL professors succeed in learning and in teaching a FL, respectively (Leons et al., 2009, p. 42).

METHODOLOGY

THE MAIN PURPOSE of this study was to delve into what teachers know about teaching FLs to students with LD and what can be done to aid both FL teachers in teaching and students with LD in learning a FL. The participants in this study were FL instructors and learning specialists from secondary schools and institutions of higher education in Maryland. A questionnaire was distributed and interviews were conducted to gather the data (see Appendices A and B). Three surveys were distributed: one for FL instructors at the secondary and college level, one for part-time faculty at the University of Maryland, Baltimore County (UMBC), and one for learning specialists at the secondary and college level.

The surveys for FL instructors in secondary schools and institutions of higher education and for UMBC part-time faculty were sent to those listed in the public domain on Maryland high school and university websites. The only difference in these two surveys was the first question, which asked for the title of the employee. Over 145 secondary education FL teachers were contacted, and thirty-six replied. One hundred and eighty-three professors, including the part-time faculty at UMBC, were contacted, and thirty-nine professors and nine UMBC part-time faculty members, including teaching assistants, responded.

A survey was sent to twenty-eight learning specialists employed at Maryland public universities, and seven responded. Out of over 200 learning specialists employed at Maryland public high schools who were contacted, sixteen responded.

The anonymous online surveys asked these instructors and learning specialists about their opinions and knowledge of FL learning for students with LD. The anonymity was a tactic to encourage them to share freely about their experiences without worry of repercussions. Links to these surveys were sent in an email to the participants because “the administration procedures of the embedded form are simpler, which leads to a higher response rate” (Dörnyei & Taguchi, 2010, p. 70). The email messages explaining this research and

requesting participation were sent multiple times to the participants because Dörnyei & Taguchi (2010) stress that “multiple attempts should be made to contact potential respondents” (71). Following Dörnyei & Taguchi (2010), all three surveys were no more than ten questions in length: “[i]n questionnaire design less is often more, because long questionnaires can become counterproductive” (12). The desire was that the participants would complete the survey with sincere, well thought-out responses.

Both the instructor and the learning specialist surveys had a low response rate due to time constraints on this research endeavor and the method of data collection. According to Dörnyei & Taguchi (2010), “despite the attractive features, online survey response rates tend to be lower than return rates of traditional postal surveys” (p. 71). The online survey tool Survey Monkey was selected for this study because it is well-known, is easy to use for both the researcher and the participant, and allows the researcher to create a survey of ten or fewer questions for free. Online surveys were considered the best option for this study because of the limited time in which data could be collected. It is faster to obtain results as they come in online rather than waiting for results to be mailed back.

In addition to the surveys, four full-time UMBC Modern Languages, Linguistics, and Intercultural Communication (MLLI) professors were interviewed. These participants were asked the same survey questions (see Appendix A) but were prompted to elaborate, to speak freely, and to add any comments or ask any questions that pertained to this topic. These participants, like the instructors and learning specialists, were kept anonymous so that they would be able to speak openly without worry of repercussions due to their opinions or experiences.

RESULTS

Full-Time Instructor Survey

OF THE SEVENTY-FIVE full-time FL instructors who participated in this survey (see Appendix A), 48% were teachers at Maryland public high schools and 52% were professors at Maryland public colleges or universities. Of these seventy-five instructors, 77.3% had knowingly taught students with LD in their FL classes. Since not all participants had consciously taught students with LD before, only

fifty-nine out of the seventy-five responded to the next question (How did you know they had a learning disability?). The participants were allowed to choose more than one response. According to the answers, 86.4% were informed that the students had learning disabilities by a note received from the school, 47.5% were informed by the student, and 16.9% received a note from a parent. Other sources of information about a student's LD included IEPs, 504 plans, a doctor's note, or the respondents' own observations of the student's behavior.

Of the fifty-nine participants who answered question four (Did you notice any of the following issues with these students?), where they were allowed to choose more than one response, the majority noticed that students with LD were frustrated (66.1%), had low confidence levels (61%), and had difficulties writing (76.3%), speaking (62.7%), listening to and understanding (57.6%), and reading (57.6%) the target language. Low grades were an issue selected by just fewer than half of the participants (49.2%), and a small minority (18.6%) shared other issues they noticed with students with LD in their FL classes. These instructors identified behavior issues and difficulty focusing in class. They also observed that students with LD needed more time than the average student to complete assignments and assessments. Another issue, according to one respondent, was the "use of translation software to get around the disability."

The fifty-nine respondents, allowed to choose more than one response, all said that they made special accommodations for the students with LD, including extended testing time (89.8%), extra help outside class (74.6%), a wide variety of in-class activities (62.7%), reduction in spelling bias (45.8%), untimed testing (33.9%), more learning resources (33.9%), and specialized homework assignments (20.3%). Some of the respondents (28.8%) explained other accommodations, such as personalized attention both in and outside of the classroom to ensure that students with LD stay on track for success, understand all assignments and concepts to the best of their ability, and have as much flexibility in the learning environment as possible.

The amount of help that the schools provided the instructors was explored next (What did the school do to help you teach these students? Please check all that apply.). The majority of this group of fifty-nine stated that the schools established an IEP for students with LD and made them aware of the situation (57.6%). For some, the schools provided no help (28.8%), while for others the schools provided training sessions (8.5%), passed on helpful literature (8.5%),

and sent teachers to conferences on this issue (5.1%). Some of the participants (33.9%) identified additional ways in which their school helped them teach students with LD. One respondent was required to have “a particular number of credits that would count as a basic level of background knowledge of Inclusion/Mainstreaming with Special Ed.” Another stated that the school “provided courses on differentiated instruction.” The elaborations of the 33.9% represented two ends of the spectrum: while some respondents complained that the school did not provide any type of substantial help to instructors, such as assistant FL teachers or special education instructors certified in FLs, others explained that their school had a great support system for students with LD and their instructors, including bringing in tutors for the students and providing courses on differentiated instruction for the teachers.

Of the seventy-five participants (Do you know what your school’s educational policies are for learning-disabled students?), 78.7% did not know what their school’s educational policies were for students with LD. The participants were then asked to share where these policies can be found and asked to choose all answers that apply. Some said that the policies were on the school website (34.7%) or in the school handbook (26.7%). Others had not seen a written copy of these policies (29.3%), while a small percentage stated that the school policies are not available (1.3%). Other places listed by participants (34.7%) included the special education student support office at the school, handouts or memos distributed to instructors, 504 plans, IEPs, and a School Improvement Plan. Some respondents stated that they were not certain where to find copies of these policies, and one respondent admitted that s/he had “never looked until I received this survey.”

Question nine asked, “What types of resources should the *school* provide *you* to address the needs of LD students? What types of resources should the *state* provide the *school* to address the needs of LD students? Please check all that apply.” A majority of the seventy-five participants responded that the following should be provided by the school: training sessions (96.1%), helpful literature (84.1%), opportunities for instructors to attend conferences which address this issue (77.4%), more learning resources for LD students (91.8%), and more teaching tools/resources (82.1%). A majority also responded that the state should fund tutoring programs (81.8%). Only five respondents did not think the school should provide any of the listed

resources; only four respondents did not think the state should provide any of the listed resources. Ten participants elaborated under the “Other” category, sharing ideas which could benefit both students and instructors. One respondent stated that it would be beneficial to “[i]nform the instructor of the student’s type of learning disability and what is detrimental to the student’s learning. All we’re told is that the student has a learning problem and needs ‘x’ accommodation.” Another respondent requested “funding for special education certified World and Classical Languages (WCL) co-teachers. WCL is treated like a core subject in that it is a requirement for graduation for almost all students, but it is not funded equally and there are no WCL co-teachers (at my school, at least).” Another idea introduced was the necessity of learning disability specialists certified in the FL area at both secondary schools and institutions of higher education. “University professors don’t have the time to go to training sessions nor to become special education teachers,” stated one of the advocates of this idea. Another respondent wrote, “disability support services should take responsibility” because “they pass off all issues to one teacher in the English department so FL issues are not handled at all.”

Of the seventy-five participants who answered question ten (Do you believe that learning-disabled students could benefit from any of the following options? Please check all that apply.), 81.3% believed that students with LD could benefit from smaller FL class sizes. Some believed that students with LD could benefit from specialized FL courses (38.7%), substitution of other courses in place of FL courses (38.7%), one-on-one time with native speakers of the target language (33.3%), and/or exemption from the FL requirement (24.0%). All participants believed that students with LD needed extra help. Other beneficial options proposed by participants (26.7%) for students with LD included having special education teachers who are certified in one or more FLs, increasing the amount and variety of student and instructor resources available, and providing tutors trained in both learning disability and FL fields.

One respondent stated, “LD students should have choices and have help in deciding which path to take. Not all learning disabilities are the same. Some students can be successful FL learners while others can’t.” Another respondent wrote, “Exempting learning-disabled students from the foreign language requirement is equivalent to telling them that they cannot learn a foreign language, which is not true and not fair to them.”

UMBC Part-Time Faculty Survey

NINE PART-TIME FL faculty members and FL teaching assistants at UMBC responded to the second survey (see Appendix A). Five of these participants were part-time faculty members and four were teaching assistants. Many (88.9%) had taught students with LD in their FL classes. Of the 88.9% who had taught students with LD (respondents could select more than one answer), 87.5% were informed that the student had a learning disability because of a note from the school, 62.5% were informed by the students, and 12.5% received a note from a parent. The participants did not list any other methods of how they found out that a student had a learning disability.

Of the eight participants who answered question four (Did you notice any of the following issues with these students? Please check all that apply.), 50% observed that students with LD had difficulty writing in the target language and 50% noticed that the students had poor grades. Respondents also noted that students with LD had low levels of confidence (37.5%), demonstrated difficulty listening to and understanding the target language (12.5%), had difficulty speaking in the target language (12.5%), and showed signs of frustration (12.5%). Some (25%) did not detect any of the listed issues, and 12.5% noticed other issues. No participants observed issues with students with LD reading in the target language. One instructor stated that students with LD “[o]ften did not complete assignments.”

Once the eight participants noticed these issues, they made accommodations for the students (respondents were allowed to choose more than one answer), such as extended testing time (87.5%), a wide variety of in-class activities (50.0%), extra help outside of class (50.0%), reduction in spelling bias (25.0%), untimed testing (12.5%), and more learning resources (12.5%). No participants made specialized homework assignments, and 12.5% said that they had made other accommodations for their students with LD, but none listed what those accommodations were.

Seven participants answered question six (a question allowing multiple responses), which addressed what the school did to help them teach these students. The majority (85.7%) said that the school provided no help. Some stated that the school established an IEP and made the teacher aware of the situation (14.3%), and 14.3% of respondents stated that the school provided means of help not

listed in the survey; however, they did not state what sort of help they received. None of the respondents were provided training by the school, sent to conferences on this issue, or given helpful literature.

Of the nine participants who answered question seven (Do you know what your school's educational policies are for learning-disabled students?), 55.6% knew what their school's educational policies were for students with LD. Of these nine, the majority (55.6%) stated that they had not seen a written copy of these policies. Some (33.3%) said that these policies could be found on the school website, while none said that they could be found in the school handbook. No respondents said that the school's policies about this were not available, and 33.3% of participants elaborated, explaining that documents with these policies are sent to the instructors. One participant said that s/he does not know where they can be found.

Question nine, which allowed multiple responses, focused on the resources that both the school and the state provided to address the needs of students with LD. The nine respondents stated that the following should be provided by the school: training sessions (100%), helpful literature (71.4%), more learning resources for LD students (66.7%), and more teaching tools/resources (66.7%). A majority (80%) responded that the state should provide funding for tutoring programs. Two people replied that teachers should be sent to conferences on this issue: one believed the school should send the teachers and the other believed the state should. One person thought that neither the state nor the school should take any of the steps listed.

When asked if students with LD could benefit from any of the options listed (multiple responses were allowed), 66.7% of the nine participants thought that these students could benefit from smaller FL class sizes. Some thought that these students could benefit from one-on-one-time with native speakers of the target language (33.3%), specialized FL courses (22.2%), or substitution of other courses in place of FL courses (11.1%). None thought that the students would benefit from exemption from the FL requirement, and no participants thought that these students did not need extra help. Many of the respondents (55.6%) shared other beneficial options for students with LD, such as helping these students complete their assignments outside of class, providing tutoring programs for them, and providing training for teachers on teaching methods which could better help these learners.

Learning Specialist Survey

TWENTY-THREE LEARNING specialists participated in the learning specialist survey (see Appendix B): 69.6% were employed at a Maryland public high school, while 30.4% were employed at a Maryland public college or university. These participants' job titles included the following: Teaching Assistant or Paraprofessional, Special Education Department Chair, Special Education Teacher, Special Educator, IEP Facilitator, Learning Specialist, Director of Disability Services, Assistant Vice President, Director of the Center for Access and Success, Learning Disabilities Specialist, Director of the UMBC Learning Resources Center, and Program Management Specialist. The specialists were asked to list up to three of the main duties of their positions, which included ensuring compliance with student IEPs, teaching, providing individualized learning support, providing tutoring to students, advocating for students with LD, and retaining students.

All of the twenty-three participants knew what their schools' educational policies were for students with LD. Allowed to choose multiple responses to the next question, a little under half of the participants knew that they could find a written copy of these policies in the school handbook (47.8%) or on the school website (47.8%). Some (13.0%) had not seen a written copy of the policies, and one said that the policies were not available. Another respondent did not know where a copy of these policies could be found, and 39.1% stated that a paper copy is distributed to every teacher or that they found copies of the policies in the special education office or on the county website.

Of the twenty-three participants who answered question six (Have you worked with learning-disabled students who were required to take foreign language classes?), the majority (65.2%) had worked with students with LD who were required to take FL classes. Nine participants elaborated about the students in the "Other" section, and three out of these nine explained that at their schools, FLs are not required for students with LD to graduate, unless they are part of one graduation path that all students can choose to follow. Some (34.8%) had not worked with students with LD who needed to take a FL and were therefore instructed to skip the next question which asked about student issues that respondents had noticed (this question allowed multiple responses).

Of the nineteen participants who had worked with students with LD required to take a FL, the majority noticed that the students had

difficulty writing in the target language and also had low levels of confidence (57.9% for each). Participants also noticed difficulty listening to and understanding the target language (52.6%), difficulty speaking in the target language (52.6%), difficulty reading in the target language (52.6%), poor grades (52.6%), and frustration (47.4%). One respondent stated that there were no signs which showed these students' learning disabilities. Some respondents (42.1%) checked "Other" and wrote of the necessity to stress requisite learning skills with students with LD, such as note taking, active listening, content area reading, organization, and study skills.

All twenty-three respondents addressed the accommodations that are made for students with LD required to take FL classes and believed that such accommodations were always made (this question also allowed multiple answers). The majority (91.3%) said that they give these students extended testing time. Some (56.5%) give extra help outside of the class. In addition, 78.3% said that teachers are made aware of the students' learning disabilities and IEPs. Other accommodations selected by participants included establishing IEPs for the students (47.8%), doing a wide variety of in-class activities (34.8%), giving them specialized homework assignments (34.8%), providing more learning resources (30.4%), providing a reduction in spelling bias (26.1%), and allowing untimed testing (21.7%). Some of the respondents (30.4%) shared other accommodations made for students with LD, such as stating that some of these students can wait to start FL classes until they are upperclassmen in order to ensure readiness for the learning process. Others shared that FL course requirements can be waived for students with LD.

This group of twenty-three also addressed the types of resources that the state should provide the school to address the needs of students with LD required to take FL classes. With multiple answers allowed, question nine revealed that the participants agreed that the state should provide the following types of resources: funding for tutoring programs (47.8%), helpful literature for instructors and administrators (39.1%), funding for school-directed instructor training (34.8%), and state-directed instructor training sessions/conferences at the schools (34.8%). Some of the participants (30.4%) listed other types of resources that the state should provide, including the suggestion that the state should give funding for FL content training of currently employed special educators. Another suggestion was to provide schools with special educators certified in the target languages. One participant suggested that the state provide

bilingual teaching assistants to read the tests aloud to students with LD who need testing accommodations. Another suggested that the FL curriculum be modified for students with LD and tailored to suit their learning needs and abilities.

The last question asked the twenty-three learning specialists if they believed that students with LD could benefit from one or more of the listed options. The options included smaller FL class sizes (65.2% agreed), specialized FL courses (56.5%), exemption from the FL requirement (43.5%), substitution of other courses in place of FL classes (39.1%), and one-on-one speaking time with native speakers of the target language (34.8%). Some of the participants (17.4%) listed additional beneficial options for students with LD, such as providing teachers with a wider variety of teaching strategies and providing schools with special educators who are certified in the target language(s).

Interviews

DURING THE INTERVIEWS, UMBC professors were asked the same questions as the full-time instructors (see Appendix A). However, these interviews were conducted in an anonymous open discussion format to allow the subjects to feel more comfortable with sharing ideas and comments. All four professors interviewed have a degree of familiarity with learning disabilities. One professor had student(s) with learning disabilities approach him/her and discuss the learning disability, two professors were notified by the school that their student had a learning disability, and one is uncertain as to whether or not s/he has had students with LD in class. Of the three professors who knew they had students with LD in class, two recognized some in-class issues such as slower writing or frustration. All three professors made accommodations for these students, such as giving extra help outside class, giving extra time for exams and assignments, increasing repetition of material, using a wide variety of in-class assignments, and reducing the standard spelling bias. Two professors described that the school's amount of support in teaching these students was minimal. One had experience teaching FLs at a university that had a unique and extensive student support program.

Two of the four professors do not know what their school's educational policies are for students with LD. One professor knows the basics of the policies, such as where to send the students for extra

help. The fourth professor is confident in his/her abilities to accommodate students with LD, but does not know exactly what the school's policies are. None of the four know exactly where to find a written copy of their school's education policy for students with LD.

In response to the question about what resources the school should provide instructors, all four professors mentioned training sessions or workshops on campus. The idea of putting together a manual of policies, resources, and examples of what has been done for certain students with LD in the past was discussed as an attainable option. Another suggestion was that the school provide instructors with relevant seminars led by learning specialists. In response to the question of what resources the state should provide the school in this subject area, instructors suggested funding tutoring programs, employing more staff in the Student Support Services office, and training FL instructors (especially new instructors).

All four professors believed that students with LD could benefit from more specialized teaching or from other options in place of the standard FL requirements depending on the type and severity of the learning disability. Such beneficial ideas included smaller FL class sizes, specialized FL courses, substitution or exemption, one-on-one speaking time with native speakers of the target language(s), and tutoring programs with tutors trained in both learning disability and FL fields.

Offering multiple discussion sections for lower-level FL classes was also suggested. These discussion sessions would be open to everyone, so that all students could benefit from the extra attention. Each section would focus on a different FL learning issue that poses problems for all FL learners, such as oral/auditory, writing/reading, grammar, and vocabulary issues.

The professors acknowledged that each option has its downsides. For example, funding might be difficult, and choices would have to be made depending on each individual learning disability. One professor stated that “[a] lot of time, LD students are not encountered in FL classes by professors because they don't even try the courses at the college level.” Also, due to confidentiality laws, unless the students with LD come forward and explain their learning disability, instructors are not aware of their specific learning issues and thus cannot help them to the best of their ability. This was described as a roadblock preventing the success of students with LD in FL classes.

DISCUSSION

STUDENTS WITH LD frequently have difficulty learning FLs. The type and severity of learning disabilities can vary greatly, making it difficult to implement general rules, protocols, or even guidelines for helping students with LD in the FL classroom. This research explored current practices in the state of Maryland and what could be done to help FL instructors and their students with LD.

The first main research question of this study explored what FL instructors and learning specialists could do to help students with LD in the state of Maryland. As seen by the instructors' comments, teachers believe that by using multiple teaching methods in the FL classroom, most learning styles will be addressed. However, the use of a variety of teaching methods is still not always enough for all students with LD due to the fact that learning disabilities can vary in type and severity. As mentioned previously, one instructor commented that the school should "[i]nform the instructor of the student's type of learning disability and what is detrimental to the student's learning. All we're told is that the student has a learning problem and needs 'x' accommodation." By identifying the specific language issues of each student with a learning disability, the instructor and learning specialist(s) can better address unique needs and tailor teaching strategies and accommodations to help the student succeed. However, confidentiality laws mean that while the school can let the teachers know that a student has a disability, the school cannot share with instructors a student's specific disability. According to both instructor and learning specialist comments, schools provide the teachers only with direction on the type of accommodations that must be made for students with LD. It is the student's decision whether or not to make the teacher aware of a specific disability. It is quite difficult for the FL instructor to help the student learn a FL without knowledge of the student's specific disability. Given this situation, teachers and students may benefit from the Universal Design for Learning (UDL), which provides teaching methods that are designed to cover all learning styles, including the learning styles of students with LD. This is a good way to be proactive and to help ensure that learning does occur and that *all* students are included in the learning process to the best of the teacher's ability. In secondary schools, this would mean using many of the recommended practices in Leons et al.'s (2009) study, such as structuring classroom activities for student success and actively

employing learning strategies in the classroom. At the college level, while some recommended practices could help, there is more flexibility for innovative and beneficial ideas, such as establishing several discussion sections per language class which each address topics on writing and reading, listening and speaking, vocabulary, or grammar. This could benefit all students, not just students with LD. Since students with LD would not be singled out, no confidentiality laws would be broken, and the success rates for all students could increase.

The second research question addressed the support that Maryland schools provide to FL instructors and learning specialists who work with students with LD in FL classes. High school teachers are provided with an IEP for students with LD to make them aware of the disability and the accommodations that are being made for the students. Learning specialists help the students by being used as supplemental aides. Both instructors and learning specialists suggested tutoring as a successful type of assistance for students with LD. Comments from the surveys suggest that in both higher education and secondary education, tutors trained in learning disability *and* FL fields could be employed by the schools to help all students. Such tutors would be well-trained in the target language(s) in addition to the educational approaches to teaching all different types of students, specifically students with LD. As was pointed out by Amend et al. (2009) and by instructors' comments from the instructor questionnaire, methods which could help students with LD could also help students without LD learn FLs better. One instructor indicated a preference for an accommodation listed in the survey and asked, "Why only for students with learning differences?" These tutoring services could be offered to all students, not just students with LD, which again would not single these students out and could increase the success rates of all learners.

Since all types and severities of learning disorders must be accommodated, another option at both the secondary and higher education levels is to hire learning specialists trained in FLs. One learning specialist made the suggestion to "[p]rovide a special educator who is certified in the language." Having a learning specialist with a high level of fluency in Spanish or French, for example, could help both the students and the instructors. The students with LD would receive the specialized help that they need, and, according to instructors' comments from the survey, FL instructors would be less obligated to find the time and the resources to try and help these students.

Just as each individual student with a learning disability is unique, each learning environment is unique. The appropriate accommodations and solutions for instructors and students will most likely vary from school to school. According to the instructors' comments from the questionnaire, it is important to establish responsibility boundaries between the instructors and the learning specialists. One instructor said, "[t]he school disavows responsibility since I am the professional!" Another instructor said, "I cannot add to my already heavy work-load yet one more responsibility [...] I don't have the time to [...] become a special education teacher." The results from the surveys show that the lines can become blurred rather easily. For example, a handful of instructors stated that their schools gave them little to no help in teaching students with LD, which made these instructors spend a large amount of their extra time researching how best to teach these students, creating customized homework assignments, and working with the students one-on-one to ensure that learning was taking place.

The third research question explored what Maryland teachers and learning specialists believe could be done at the state level to support schools. If school administrators feel as though the instructors should take charge of most of the responsibilities related to their students, then the state-run school systems should consider creating and obtaining funding for workshops for FL instructors or sending FL instructors to school-run or state-run conferences, which review learning disabilities, teaching methods, accommodations, and more. The results of the instructor survey suggest that instructors are supportive of such ideas. An instructor referred to a specific experience with a student with a learning disability, stating, "[i]t would have been important to have been trained or advised about the topic." One FL instructor from the survey pointed out that s/he is not a learning specialist and does not have the time to take classes or figure out how to teach students with LD. Workshops and conferences would not take up too much of an instructor's time, would add to his/her knowledge of learning disabilities, and would suggest new methods of teaching FLs to students with LD. These suggestions could be applied to help increase FL learning among students with LD, to assist FL instructors in teaching these learners, and to assist school systems in guiding instructors and learning specialists in this area.

The results of this study demonstrate the need to increase knowledge of how to best help students with LD. The results of the questions that asked if the instructors knew their schools' education-

al policies for students with LD and where a written copy of these policies could be found were quite alarming. Many did not know the schools' policies or the location of the policies. It is federal law to successfully accommodate students with LD in high school and college (utilizing IDEA and ADA regulations). Since the results revealed that there are many instructors who do not know their schools' policies for students with LD, this indicates the need for schools to increase teachers' awareness of students with LD and to stress the importance of accommodating these students according to school policies and federal law.

CONCLUSION

THE PURPOSE OF this study was to explore FL learning for students with LD and to see if suggestions could be made to further improve their success in learning a FL. A variety of options and pathways should be made available to students and instructors. The suggestions discussed above could contribute to higher success rates for students with LD in FL classes. Of the suggested methods, there are three possible solutions that could most easily be implemented to begin to better address the needs of students with LD. The first possible solution is the creation of a handbook or guide for schools. Since no two students with LD will have the same issues in the FL classroom, a generic handbook or protocol is difficult to implement in schools. However, respondents recommended a handbook featuring information about learning disabilities, FL teaching methods, and examples of previous accommodations made at each school along with information about learning disabilities. This handbook could serve as a guide for learning specialists and instructors in regards to what options could be made available to students with LD in FL classes. The state could create a handbook to send to all schools, and each school could also make a handbook with practices that have been put in place for students with LD to help them succeed. The second possible solution is holding workshops and/or conferences for FL instructors that specifically address FL teaching and learning methods for students with LD. The third possibility is hiring special educators who are certified in one or more of the target languages taught at each school. Such educators already exist, but they are few in number.

Despite the insights this research has provided, the study has several limitations. First, the online questionnaires were completed

voluntarily by a small number of instructors and learning specialists within the state of Maryland. The survey was short and the sample is not large enough to truly be representative of all FL instructors and learning specialists. Nonetheless, the sample is large enough to conclude that the experience of FL learning for students with LD does vary greatly among learners, among instructors, and among schools. Another limitation is the fact that not all instructors were distinguished between part-time, full-time, or teaching assistant status due to insufficient information on each school's public domain website. Knowing the participants' statuses could have been helpful to this study.

This study represents a small portion of what future studies could do in this subject area. With more time and funding, this study could continue to survey and interview instructors and learning specialists across the state of Maryland in order to gather more data and to share more personal experiences and opinions of FL instructors and learning specialists. This work will hopefully raise awareness of this issue and prompt others to delve into research of their own that could further aid students with LD in FL classes.

APPENDIX A: INSTRUCTOR SURVEY QUESTIONS

Key:

FT = Full-time secondary and higher education instructors questionnaire

PT = Part-time higher education instructors questionnaire

Note: When I began this study, I referred to students with LD as “LD students.” Thus, the term “LD students” was used in the questionnaires.

1) Please choose which statement applies to you:

FT: I am a foreign language teacher at a Maryland public high school; I am a foreign language professor at a Maryland public college or university

PT: I am a part-time faculty member at a Maryland public college or university; I am a teaching assistant at a Maryland public college or university; I am a conversation hour leader at a Maryland public college or university

2) Yes or No: Have you taught learning-disabled students in your foreign language classes? (If the answer is “no”, please skip to Question #7.)

3) How did you know they had a learning disability? Please check all that apply.

School note

Parent note

The student(s) told you
Other (please specify)

4) Did you notice any of the following issues with these students?

Please check all that apply.

Difficulty listening to and understanding the target language
Difficulty speaking in the target language
Difficulty writing in the target language
Difficulty reading in the target language
Low levels of confidence
Frustration
Poor grades
None of the above
Other (please specify)

5) What accommodations did you make for these students? Please check all that apply.

Extended testing time
Untimed testing
More learning resources
Wide variety of in-class activities
Specialized homework assignments
Reduction in spelling bias
Extra help outside class
No special accommodations are made
Other (please specify)

6) What did the school do to help you teach these students? Please check all that apply.

Provided training session(s)
Sent you to conference(s)
Passed on helpful literature
Established an IEP for the student(s) and made you aware of the situation
The school did not help me
Other (please specify)

7) Yes or No: Do you know what your school's educational policies are for learning-disabled students?

8) Where can a written copy of these policies be found? Please check all that apply.

School website
School handbook
I have not seen a written copy of these policies

School policies are not available

Other (please specify)

9) **School vs. State:** What types of resources should the *school* provide *you* to address the needs of learning-disabled students? What types of resources should the *state* provide the *school* to address the needs of learning-disabled students? Please check all that apply.

Training sessions

Helpful literature

Send you to conferences which address this issue

More learning resources for learning-disabled students

More teaching tools/resources

Funding for tutoring programs

None of the above

Other (please specify)

10) Do you believe that learning-disabled students could benefit from any of the following options? Please check all that apply.

Smaller foreign language class sizes

Specialized foreign language courses

Exemption from the foreign language requirement

Substitution of other courses in place of the foreign language classes

One-on-one speaking time with native speakers of the target language

Learning-disabled students do not need extra help

Other (please specify)

APPENDIX B: LEARNING SPECIALIST SURVEY QUESTIONS

1) Please choose which statement applies to you:

I am employed at a Maryland public high school

I am employed at a Maryland public college or university

2) What is your job title?

3) Please list up to three main duties of your position.

4) Yes or No: Do you know what your school's educational policies are for learning-disabled students?

5) Where can a written copy of these policies be found? (Multiple responses allowed.)

School website

- School handbook
- I have not seen a written copy of these policies
- School policies are not available
- I do not know
- Other (please specify)

- 6) **Yes or No: Have you worked with learning-disabled students who were required to take foreign language classes? (If the answer is “no”, please skip to Question #8.)**
- 7) **Did these learning-disabled students show signs of any of the following issues in their foreign language classes? Please check all that apply.**
- Difficulty listening to and understanding the target language
 - Difficulty speaking in the target language
 - Difficulty writing in the target language
 - Difficulty reading in the target language
 - Low levels of confidence
 - Frustration
 - Poor grades
 - There were no signs that showed these students’ learning disabilities
 - Other (please specify)
- 8) **What accommodations are usually made for learning-disabled students who are required to take foreign language classes? Please check all that apply.**
- Extended testing time
 - Untimed testing
 - More learning resources
 - Wide variety of in-class activities
 - Specialized homework assignments
 - Reduction in spelling bias
 - Extra help outside class
 - Establishment of an IEP
 - Teachers are made aware of their learning disability and IEP
 - No special accommodations are made
 - Other (please specify)
- 9) **What types of resources should the state provide the school to address the needs of learning-disabled students who are required to take foreign language classes? Please check all that apply.**
- Funding for school-directed instructor training
 - Helpful literature for instructors and administrators
 - Funding for tutoring programs

Hold state-directed instructor training sessions/conferences at the school
Other (please specify)

10) Do you believe that learning-disabled students could benefit from any of the following options? Please check all that apply.

Smaller foreign language class sizes

Specialized foreign language courses

Exemption from the foreign language requirement

Substitution of other courses in place of the foreign language classes

One-on-one speaking time with native speakers of the target language

Other (please specify)

REFERENCES



- ADA Home Page*. (n.d.). Retrieved from <http://www.ada.gov/>
- Amend, A. E., Whitney, C. A., Messuri, A. T., & Furukawa, H.** (2009). A modified Spanish sequence for students with language-based learning disabilities. *Foreign Language Annals*, 42(1), 27-41.
- CAST: About UDL*. (n.d.). Retrieved from <http://www.cast.org/udl/>
- Crombie, M., & Schneider, E.** (2003). *Dyslexia and foreign language learning*. London: David Fulton Publishers.
- Dörnyei, Z., & Taguchi, T.** (2010). *Questionnaires in second language research: Construction, administration, and processing*. New York: Routledge.
- IDEA 2004*. (n.d.) Retrieved from <http://www.ldonline.org/features/idea2004>
- Learning Disabilities: An Overview*. (n.d.) Retrieved from <http://www.ldonline.org/article/5613/>
- Leons, E., Herbert, C., & Gobbo, K.** (2009). Students with LD and AD/HD in the foreign language classroom: Supporting students and instructors. *Foreign Language Annals*, 42(1), 42-54.
- Mastropieri, M., & Scruggs, T.** (1990). Teaching abstract vocabulary with the keyword method: Effects on recall and comprehension. *Journal of Learning Disabilities*, 23(2), 92-96.
- Sparks, R.** (2009a). At-risk second language learners: Problems, solutions, and challenges. *Foreign Language Annals*, 42(1), 3-4.
- Sparks, R.** (2009b). If you don't know where you're going, you'll wind up somewhere else: The case of 'foreign language learning disability'. *Foreign Language Annals*, 42(1), 7-26.
- Young, D. J.** (2009). Working toward shared visions of successful language learners. *Foreign Language Annals*, 42(1), 5-6.

A DIGITAL SKELETON KEY TO ART

Symbolism of Light and Dark in European Oil Painting

john winder

John Winder will graduate in May 2013 with a major in computer science and a minor in art history. As an Honors College student, John strongly believes that studying the arts and humanities is important and that interaction across disciplines is useful, enlightening, and essential. After graduation, he plans to go to graduate school where he hopes to further investigate how computer science might aid art historical scholarship. His interdisciplinary research received a UMBC Undergraduate Research Award. He would like to thank the UMBC Office of Undergraduate Education, the editors of the *UMBC Review*, and his mentor Dr. Preminda Jacob for her encouragement, enthusiasm, and guidance. He would also like to thank his parents and brother for their enduring support.



Inspired by an aside in a math textbook, I realized that digital images contain more information than their surface appearance, that this information can be unlocked, affording new perspectives while remaining true to the original, and that an image's regions of light and dark constitute the key to seeing the image in new ways. For digital images, the values of light and dark can be readily enhanced and abstracted, facilitating qualitative and quantitative analyses. For many oil paintings, the contrast of light and dark is crucial to the work's symbolic meaning and its structural form. An algorithm that abstracts and enhances a digital image in a light-dark context, then, when applied to an image of an oil painting, would help elucidate how chiaroscuro (the technique of shading) was used by oil painters as a narrative and compositional tool. This research both discusses the history and applications of chiaroscuro and presents a computer program that deconstructs digital images to aid art historians and laymen alike in the investigation of light and dark in oil paintings.

INTRODUCTION

CHIAROSCURO, FROM THE Italian *chiaro* (clear, light) and *oscuro* (dark), is the intentional use of contrasting light and dark, or shading, in artwork to give the appearance of three-dimensional form. The presence of chiaroscuro in paintings, then, is the depiction of shadows to achieve realism. Chiaroscuro not only permits mimesis in the painter's work but also infuses the art with symbolic gravity. In particular, the oil painters of Europe from 1500 to 1900 CE endeavored to present an illusion of reality in their artwork, making extensive and sometimes exaggerated use of chiaroscuro.

The phenomenon of the shadow is one of absence, as it exists only where light is not. Across cultures, it is associated with varied mystical aspects, including unreality, devilry, and the concept of the soul (Gombrich 17). Further, the shadow is an ancient metaphor for both the art of painting itself and our ability to understand the universe (such as the fire-cast shadows from the Allegory of the Cave in Plato's *Republic*). The painters who used chiaroscuro and struggled to mimic shadows in paint knew the ancient and spiritual connotations of light and dark. Regarding Leonardo da Vinci's predecessors and contemporaries, who did not use chiaroscuro, Ernst Gombrich remarked:

We soon realise that some of the greatest observers of nature appear to have deliberately avoided the cast shadow. However rich their palette and their mastery of tone and colour, they show us a shadowless world. It looks indeed as if many of these masters had studiously avoided inserting such shadows, as if they regarded them as a disturbing and distracting element in an otherwise coherent and harmonious composition. (19)

The implication, therefore, is that when painters *did* use chiaroscuro, the shadows of those shadowed worlds possessed the power to disturb and command our attention. Indeed, by examining the painter's controlled and deliberate use of chiaroscuro, we gain insight into the crucial role of the shadow (and the light necessary for its existence) as a compositional and allegorical tool in European paintings.

PROJECT GOALS

THE INTENTION OF “A Digital Skeleton Key to Art” is twofold: to discuss the history and implications of chiaroscuro and to create an algorithm to aid in the investigation of its significance in both the symbolism and realism of European oil paintings. The latter task deconstructs, using a computer program, a painting’s regions of light and dark to understand their structure and function in terms of composition and figurative content. In the resultant digital analysis of a painting, we are given a novel perspective for viewing the artwork purely and solely in the context of its own brightness and darkness, a feat impossible with an unaided eye. Thus, with the assistance of a computer, a step-by-step procedure for the excavation of hidden aspects of chiaroscuro in works of art makes available an analytic tool for the art historian.

This analysis of digital images is accomplished in two ways: first, by abstracting a painting into regions of light and dark to understand composition better; second, by enhancing the contrast of an image to examine detail more closely, to discuss content better, and to consider the conventional assumptions about a work’s narrative in light of this new perspective. Together with the historical discussion of chiaroscuro, the final product of this project is an algorithm and an implementation of that algorithm in the MATLAB programming language. The program allows anyone with a digital image to examine it in a few different ways by abstracting and enhancing the image and by providing basic statistics so that the relative brightness can be compared qualitatively and quantitatively among paintings.

RATIONALE

KEY TO SIGHT, and thus to the perception of reality, are the gradations between light and dark, between the binary opposition of white and black, and among the intermediate spectrum of grays which allow the eye to distinguish objects, depth, and form. As paintings are an almost entirely visual experience, they intrinsically depend upon light and dark. To painters who create representational works that strive for verisimilitude, there is added pressure to exploit and heighten the contrasts between light and dark to achieve a greater semblance of reality. Paintings created in this manner contain more

information, both physical and symbolic, than can be seen by a cursory or unaided glance.

This algorithm facilitates the perusal of such paintings by providing a tool for the art historical investigation of the formal and contextual roles of light and dark in those works. The disciplinary practice of art history was made more feasible by photographic reproductions so that scholars did not require direct contact with the original objects of their study. In this century, we possess higher quality and more faithful photographic reproductions than ever before. Moreover, digital images actually grant us the ability to interact with the information stored in the photographs. This advancement enables us to see more in, and do more with, the paintings than we could by merely viewing a photographic reproduction in a book.

The computer program is designed to facilitate the work of the art historian and layman alike. Anyone can explore its capabilities, investigate paintings through it, and consider the relationships among the painting's form, content, light, and dark. The ultimate goal is to enhance future interpretations of these painted works constituting the world's heritage, even if just by encouraging the art aficionado to see more than what is plainly visible to the unaided eye.

PRACTICAL SIGNIFICANCE OF THE SHADOW IN ART

IN THE SPACES that exceed the reach of light, in recesses, cracks, and holes, along curved planes and textured surfaces, and behind and around illuminated objects can be found the presence of a shadow. Even commonplace shadows may be dramatic or subtle, like the shadow under the door to an unlit room or the shadow of the nose across a face. Though omnipresent, shadows are still rarely appropriately recognized as bestowing the objects around them with dimensionality. Attempts to reconstruct reality, whether digitally or with paint, rely upon shadow as much as color and line (Cavanagh 395). Hence, visual cognition of depth, form, texture, and the spatial relationships between objects (that is, reality) depends upon the presence of shadows for human perception. Any artist who wishes to paint representational images, then, must study shadows.

Mindful observation of shadows became a common practice among Western artists during the Renaissance, concurrent with the introduction of oil paint as the standard medium (Stoichita 48-49).

The goal of mimesis, authentic portrayals of form and light, necessitated an intense study of shadows among painters from 1500 to 1900 CE. The extensive use of shadows and shading seen from the time of the Renaissance through Romanticism was essential for paintings that aspired to allegorical realism, since any observation of reality must admit the ubiquity of shadows. Shading heightened the illusions of depth and dimensionality on the surface of a painting. Hence, shadowed scenes are endowed with their own personal lighting and semblance of space. Impressed upon the viewer is a sense that the objects, figures, and places truly exist, appearing as they would if viewed through a window. Victor Stoichita notes the gradual shift of the Italian artists from shadowless to shadowed worlds by discussing Giovanni di Paolo's *The Flight into Egypt* (1436), in which *some* of the buildings, people, and plants cast shadows while, interestingly, the primary figures do not (46-48).

Artists attentively constructed paintings so that the composition of light and dark would achieve depth and form. The development of this stylistic technique, known as chiaroscuro, led to paintings with regions that recede into blackness, while protrusions and surfaces hit by light are lighter. The constructed nature of paintings themselves, combined with the inclination of art to recount a story, naturally allowed the artists to incorporate symbolism that was spiritually, emotionally, or simply aesthetically significant.

SYMBOLISM OF THE SHADOW IN ART

FOR A VERY long time, the shadow has been a metaphor for painting itself. An ancient tale explains this indelible connection: a potter's daughter dreads the approaching date of her beloved's departure. He is leaving for war; their parting may well be the last moment they share together. In a fit of inspiration tinged with anticipated heartache, the young woman creates a unique memento of her love: she traces the silhouette of his shadow on the wall of her home. According to some, her father, the potter, uses the recorded shadow to construct a clay figure in the semblance of the man. The beloved soldier loses his life, perhaps heroically. All that remains is his frozen shadow, as perceived by his love, and the totem or doll that her father has created.

This story of the Corinthian maid, and its variations, are cited collectively by ancient historians as the apocryphal origin of painting

(and sometimes sculpture) in the West. It is the pseudo-historical account the Greeks used to explain the invention of their visual arts (and to distract from their sizable artistic inheritance from the Minoan and Egyptian civilizations) (Stoichita 11-20). The conditions of the story are appreciably melodramatic; art is born from a scenario of trauma, love, and beauty, ending with mortal loss. The crux of the legend is that all of these emotions are tied to and epitomized by the recorded shadow, art. In this way, the shadow is a metaphor for painting and its origin and carries all of the other connotations, good and evil.

A HISTORY OF CHIAROSCURO

ALTHOUGH THE GREEKS did not invent painting, as their myth might claim, an ancient Athenian was the first known artist to develop the technique of *skiagraphia*, the painting of shadows. First seen in the High Classical era, roughly 450 to 400 BCE, *skiagraphia* (equivalent to the modern notion of *chiaroscuro*) is credited to the painter Apollodorus, a mentor of Zeuxis. Apollodorus and his contemporaries helped move Greek painting from the linear style typical of vase painting to more elaborate techniques of illusionism, primarily for flat surfaces such as walls (“Painting, Western”). Though very few paintings from antiquity are extant, historical accounts from the time make mention of significant stylistic changes. One such record, by the rhetorician Quintilian, notes that paintings from earlier eras could be recognized by their flat, or solid, colors (Keuls 7). Shadows afforded dynamic color and thus a heightened sense of realism in a painted surface. One of the only ancient Greek paintings still extant, fortunately exhibiting *skiagraphia*, is *Hades Abducting Persephone* (see fig. 1), a painting on the wall of a tomb in Vergina, Macedonia (“Painting, Western”). The red robe, an impressive study of shadows in the folds of an agitated cloth, adds depth and a sense of motion to the painting. Any impression of fidelity to reality inspired by this work would be marred without the artist’s observation and incorporation of shadows.

Skiagraphia, a tool at the hand of skilled painters, breathes life into paintings, just as the shadows in the ruffled folds of the cloth in the Vergina painting inspire a sense of a violent rush. High Classical era painters no longer painted objects with a single, unbroken color (Keuls 7, 12-13). The use of *skiagraphia* unlocked realism in paint-

ing: the ability to depict believable three-dimensional forms made the portrayals of the figures' stories more immediate and powerful. Thus, to enhance their paintings, artists aspired to couple sophisticated observation with intentional manipulation of light and dark to create a convincing yet aesthetically appealing semblance of reality.

GREAT PRACTITIONERS OF CHIAROSCURO

AN ARRAY OF great artists, including Leonardo, Caravaggio, Rembrandt, David, and Goya, owe their popularity in no small part to their command of chiaroscuro. Paintings by these artists depict glowing colors and lights, both forceful and gentle, which shine on figures to suggest hope or salvation, and shadows that, like dark clouds, shroud other figures in gloom. Beginning in the sixteenth century, Leonardo was the first artist to devote extensive study to light and, in particular, the nature of shadows (Baxandall 151-155, "Chiaroscuro"). Sfumato, the smoky, gentle blending of shadows, is the aspect of chiaroscuro for which he is so famous and which is exhibited matchlessly in the *Mona Lisa*.

The subtle colors of Leonardo gave way to the gloom of tenebrism, prevalent in the counter-reformation of the seventeenth century. Tenebrism was characterized by extensive use of darkness and shadows and was inaugurated by Caravaggio with his supremely theatrical use of chiaroscuro. In Caravaggio's paintings, figures nearly always appear as though hit by a spotlight, and yet they are surrounded by shadows without fail. For instance, in his *Supper at Emmaus* (see fig. 2), our eyes are treated to a contradictory scene where an intense light shining from above illuminates all the figures, while the room and the spaces below and along the sides are utterly, hopelessly black. Further, Caravaggio used chiaroscuro compositionally. This is seen in the outstretched hand of Jesus, the focal point, which is emphasized through its placement at the joint where the left arm of the standing figure meets the shadow cast against the wall. The Baroque painters who followed Caravaggio were indebted to him for his innovative use of light (Bell 141-142).

While less illuminated than Caravaggio's paintings, the work of Rembrandt still exhibits a self-evident mastery of light. Rembrandt displayed an intense awareness of chiaroscuro even at an early age, as is evidenced by the mysterious shadow across his face in his self-portrait (see fig. 3). The light in his works inspires a sense of humbling

harmony; one could not easily find such an unsettling use of chiaroscuro in his corpus.

Conversely, David, who worked in the eighteenth century in the neoclassical style, used lighting and shading in a style reminiscent of Caravaggio's, where figures seem like thespians on stage, framed against a dark backdrop. His use of shadow was melodramatic, as in *The Lictors Bring to Brutus the Bodies of His Sons* (see fig. 4). The scene truly evokes the feeling of an ancient Greek tragedy. Vivid above and beyond the head of Brutus, who sits cloaked in a shadow, is the lifeless body of one of his sons, recently executed for treason by order of Brutus himself. The darkness in which Brutus broods embodies the quintessential sinister and unwholesome connotations of the shadow.

Lastly, in the early nineteenth century, Goya created a haunting protest against the futility of war. In his *The Third of May, 1808*, a firing squad is seconds away from slaughtering a rabble of captive men, having already fired once. One of the victims, clad in a brilliantly white shirt, throws his arms out passionately in a sign of surrender. He faces the dusky, blurred mass of soldiers, under a pitch-black, moonless sky. There is no hope in this scene. Goya used light to instill this man with a defiant innocence, one patently and absolutely helpless. Chiaroscuro alone makes the anti-war message of *The Third of May* possible.

Throughout the history of chiaroscuro, shadows and light together enabled artists to create works that have an authentic semblance of reality, that convey ideas through symbols, and that forever leave their viewers with an unforgettable image.

BASIS OF THE ALGORITHM – “A DIGITAL SKELETON KEY TO ART”

THE ALGORITHM OF “A Digital Skeleton Key to Art” is a way of exploring a digital image. A bitmap or raster image (the most common data structure for digital images) stores its information in a matrix of pixels, where each pixel is a cell that specifies a color. Although different types of image formats store it differently, color is typically represented by three integers that act as coordinates in a color space. Frequently used is the Red-Green-Blue (RGB) space, where the first integer represents the degree of red, the second the degree of green, and the third the degree of blue in a single color pixel. RGB provides a good approximation of the colors that are able to be assembled and interpreted in the brain: red, green, and blue

wavelengths of light are the only ones the cones of the retina can sense (Fairchild 8-9). Thus, with any pixel able to represent virtually any color, a digital image of a painting can be thought of as a faithful facsimile of the original.

In order to abstract images into regions of light and dark, it was necessary to isolate the most relevant feature of an image: its luminance. Luminance, also called value, is the measure of a pixel's brightness along the spectrum from black to white. A grayscale image consists of pixels that have just one integer representing the pixel's brightness. Most commonly, luminance values are stored as a byte and consequently range from 0 to 255, from total black to total white respectively (with a gradation of colors between). Where the RGB color space is a three-dimensional system, luminance is a one-dimensional number line, allowing for a simpler relationship between neighboring pixels (as there is only one shared trait to consider instead of three), which is crucial for how the algorithm abstracts luminance. Though color *does* influence perceived brightness, it is possible to compensate those effects when converting an RGB image to grayscale (as is the case with MATLAB). Since luminance, and not color, is the sole trait that needs to be examined when considering light and dark, the digital images to be analyzed are first converted to grayscale.

Conversion to grayscale as a matrix where each element is a single integer between 0 and 255 allows a painting to be expressed mathematically. Mathematical operations can then be performed on this manifestation of the painting to manipulate its appearance. For instance, it is possible to blend the luminance values of neighboring pixels by averaging the value of a pixel with its neighbors. Specifying an arbitrary size of square submatrices and averaging all elements within each submatrix to a single value produces abstraction that yields regions of homogeneous, intermediate luminance. This abstraction could be seen as a way of digitally stepping back from the painting, considering a more blurred version, to observe the painting's structure of light and dark.

Abstracting a painting into regions of light and dark removes detail while retaining the compositional arrangement, letting us examine the work wholly in the context of its luminance. Using this phenomenon as an analogy for studying language, the linguist Peter Culicover considers how, when you stand back from the blocky painting of *Gala* by Dalí, *Gala Contemplating the Mediterranean Sea Which at Twenty Meters Becomes the Portrait of Abraham Lin-*

coln – Homage to Rothko (second version), it appears to be a profile of Abraham Lincoln. Of this blocky metaimage, which is equivalent to an abstracted version of a painting produced by the final algorithm, Culicover says, “we are able to see the image by eliminating detail, in fact, a considerable amount of detail that is not essential to the image of Lincoln” (2). The structure that Dalí playfully and deliberately inserted into his painting, the portrait of Lincoln, arises as we remove detail. Essentially, Lincoln’s head *is* the composition of the painting. If we regard the paintings of Europe from 1500 to 1900 CE in the same way, stepping back and considering the abstracted version, we can see their composition emerge as detail is removed. This image, then, is the light-dark skeleton of the painting.

This kind of abstraction is the key to removing detail while retaining structure and the relationship of light and dark. An algorithm describing how to perform such an abstraction gives researchers the power to inspect the chiaroscuro of paintings within this light-dark context. This process is not divorced from the human observer and is not intended to supplant the ability of the mind to make connections and associations. Rather, such an algorithm supplements the ways in which art historians, or laypersons, analyze the composition of a painting, with particular regard to its chiaroscuro, its regions of light and dark in relation to each other. By presenting the painting in this novel and abstract context, the viewer is more informed of the significance of shadow as a compositional, and perhaps symbolic, element. A further consequence is that the abstracted images can be scrutinized not just qualitatively but also statistically. By opening paintings to quantitative analysis, the algorithm facilitates the forming of connections between light and dark and the painting’s meaning. Without the algorithm, these connections may not have been possible or as readily apparent. Thus, a step-by-step procedure explaining this abstraction constitutes a *skeleton key* that unlocks the light and dark of a painting.

DEVELOPMENT OF THE ALGORITHM

THE ORIGINAL INTENT was for the algorithm to be theoretical, leaving the codification of the algorithm outside the scope of this project, with examples created by executing the algorithm by hand. However, a problem immediately encountered was the realization that, even in a small image, there are a huge number of individual pixels. For instance, in an image of 200x200 pixels there are 40,000 unique pixels, which are far too many to work with consistently and

accurately without the mechanized hand of the computer. This project called for even higher resolution images. Therefore, the conclusion was that a program implementing the algorithm was necessary so that at least part of the algorithm could be automated. MATLAB was selected for this project because it is designed for manipulation of matrices and includes image processing tools which helped speed the development process.

Initially, the algorithm existed only as a procedure: a series of steps to take a digital image of a painting, convert it to grayscale, overlay grid upon it, and abstract each cell of the grid to a single, average color (that is, each grid cell is a square submatrix of the image). Two initial mockups of before-and-after-images, created in an image editor, are examples of how the *unlocked*, or *skeleton key*, abstract versions of paintings produced via the algorithm would look (see fig. 5).

After a few drafts, the algorithm was more refined but still rough (see fig. 6). This procedure describes the thought process comprising both the abstraction and analysis of the works of art. The first six steps, however, are mechanical and thus can be automated by the computer. Therefore, extending the reach of this project from its initial aims, a simple program was written in MATLAB to implement the algorithm.

Referencing *Digital Image Processing using MATLAB*, by Gonzalez, Woods, and Eddins, and the documentation provided in MATLAB's image processing toolbox helped elucidate what operations would be necessary to code the algorithm. MATLAB afforded relatively easy manipulation of digital images of paintings and treated images in the exact way described above (specifically, using pixel data as numbers in a matrix). A rudimentary program was eventually coded, omitting the presence of a grid because its black lines would make the image appear darker. Named *skeletonkey*, the MATLAB function (program) written for this project loads an image, abstracts it by a specified amount, and generates the pixelated, abstracted version. Over time, the *skeletonkey* function was augmented to include more features and even a variation (see fig. 7 for the most recent version of the function, in MATLAB code).

Due to the built-in libraries of MATLAB, it was fairly straightforward to incorporate into the *skeletonkey* program the alternative to the abstracted images, namely, the *enhanced* images. These images are created by heightening the contrast of light and dark to reveal and highlight detail, thus achieving a complementary perspective to the abstracted images. Furthermore, a variant of the program was made

called progressive, which creates successively more abstract images simultaneously and stores them into an animated .gif file. This .gif shows frame-by-frame the evolution of the painting as it undergoes abstraction. As was speculated, this program best demonstrates how detail is gradually removed when successive abstraction is applied to a painting, yet the overall structure and light-dark form are retained.

PROGRAM SPECIFICS – ABSTRACTING IMAGES

BEFORE INITIATING THE program, it is necessary to obtain a digital photograph of a painting that is faithful to the original. Applying the algorithm to a high-quality digital photograph of a painting involves a series of steps. There are two requirements to the function: the filename of the digital image and the integer amount by which to abstract it (essentially, this number is the length of the diagonal of each block in the final image). First, the image is set to grayscale. Second, the enhanced image is created using MATLAB's `adapthisteq` function, which is discussed later. Then, after storing the dimensions of the file, the function iterates over the elements of the image in submatrices whose sizes are determined by the second parameter. Drawing a grid, as specified in the algorithm, is not explicitly necessary since dividing the matrix into submatrices achieves the same effect. The values of the pixels in each submatrix are averaged together, and all are set to that new value. Thus, the submatrix becomes a collection of pixels with a homogeneous value. This averaged submatrix replaces the corresponding submatrix in the image. After all submatrices are replaced, we have the final, "pixelated" image. The last step is to have MATLAB create a histogram of the pixel values for each image. The products of the program are two new images (an abstracted image referred to as the skeletonkey version and an enhanced one referred to as the `adapthisteq` version) and their respective histograms.

As discussed above, the purpose of abstracting the image is to aid our recognition of which areas of the image's composition are generally lighter or darker. Consider, for instance, a lone white pixel in a field of black. Although such a pixel might be the single lightest *spot*, it would not be read in the eye of the viewer as a bright *region* of the painting. It might be a speck on the painting's surface or an artifact from the photographic process, and thus not truly part

of the painting. In the blocky image created by the algorithm, our eyes can more readily identify how regions of light and dark affect the composition of the work. Moreover, the histogram provided by the program allows for a quantitative consideration of the degrees of light and dark in a painting. Since each light-dark value is given on a scale from zero to 255, from black to white respectively, we can thus empirically say what is darkest and what is brightest within a given painting. For any image we examine, when more pixels are in the higher number values (that is, more pixels are closer to 255 or pure white than to zero), it is a lighter painting. Likewise, it is a darker painting when there are more pixels near zero. Such graphical representations of the data of a painting facilitate the comparison of light and dark among paintings and provide a new way to consider the chiaroscuro of any given work.

PROGRAM SPECIFICS – ENHANCING IMAGES

TO CREATE ENHANCED images or images where contrast is heightened to reveal more detail, there is a convenient function built into MATLAB called `adapthisteq`, which stands for adaptive histogram equalization. This function takes small regions within the image and alters the pixel values to balance the histogram of that region of the image. The result is that contrast is heightened and that small variances in contrasts are made more noticeable (whereas they would not be visible by similar functions that considered the image globally). For instance, if we compare the two images of Franciabigio's painting (see fig. 8) and examine the grayscale version, where MATLAB has performed the `adapthisteq` function, the difference from the original is stark, exemplifying the consequence of chiaroscuro to defining depth in a scene. Many details, hidden by the murkiness of the painting itself (it has also darkened over time), are made considerably clearer. The whole background is revealed, the crown on the step in the lower right is plainly visible, and the dove representing the Holy Ghost can be seen overhead, with rays of light shooting out from its body.

The details enhanced by `adapthisteq` can be surprising. For instance, the `adapthisteq` function uncovers an interesting detail in Caravaggio's *Ecce Homo* (see fig. 9). As with many Caravaggio paintings, the background recedes into an impenetrable well of darkness. However, in *Ecce Homo* the darkness is interrupted by an almost

imperceptible band of light: a halo graces the head of Jesus and is not perceivable to the unaided eye, except for a portion just above his ear. The *adapthisteq* function reveals the halo in its full form, standing out vividly. The crown of thorns and his ear are also illuminated. It is astounding to see aspects of the painting that were already present but nearly unnoticeable before.

Caravaggio's style incorporates regions of utter darkness and people who seem bounded in a mortal coil. These notions may, at first, seem inverted by the halo. The halo is vestigial, an artistic remnant from a time when holy men were painted with otherworldly distance. Unlike them, Caravaggio's Christ in *Ecce Homo* has a physical presence. Perhaps the artist purposefully dimmed the halo, signifying the bleakest time in the life of Jesus, when Pilate announces, "Ecce Homo," *Behold the man*, as the torturer puts a cloak about the recently-scourged Christ. Yet that wispy veneer of paint is concrete in the digital copy of the painting. It is a hidden light, almost engulfed in darkness, but still present. Its existence mirrors the downcast gaze of Christ, humble and dim.

The *adapthisteq* function effectively dramatizes the *chiaroscuro* of the painting, helping to reveal aspects already present in the work. Perhaps more profoundly, the digital process *adapthisteq* enhances detail strictly based on relative value of light and dark, presenting a version of the painting that could not be seen with the naked eye. As opposed to similar, less sophisticated algorithms in digital image editors, *adapthisteq* intelligently raises the contrast of the image by adaptively matching a new, equalized histogram to the initial one. The result is a generated image that remains true to the original yet reveals details more noticeably based on both their location in the painting and their light-dark values (again, this is something other image editors could not achieve). While *adapthisteq* may not reveal new details to a viewer with a trained eye, it can still provide everyone with new information about those details by highlighting them within the specific context of light and dark. The *skeletonkey* function, as discussed above, does the reverse by removing detail through abstraction in favor of elucidating the interactions of light and dark. With both *adapthisteq* and the initial *skeletonkey* as parts of the algorithm, the significance of *chiaroscuro* in European oil paintings can be unlocked.

PROGRAM APPLICATION FOR THEMATIC ANALYSES OF OIL PAINTINGS

THIS PROGRAM CAN be used to consider a group of oil paintings, related by some common theme, and investigate questions related to their light and dark elements, examining how artists handle the chiaroscuro in similar contexts. Such an analysis might be called a vignette. The following is a sample vignette that investigates the significance of light in religious paintings of the Late Renaissance.

Paintings of Jesus after his death are among the most fascinating religious European oil paintings because they present a scene from the Christian canon when the first followers of Christ felt utterly hopeless. The dead Christ is the artist's opportunity to paint a moment of profound spiritual anguish, to paint a dead body that is at once seemingly mortal and yet divine. The subject of the lamentation rivals all other religious scenes as perhaps the most psychologically distressing.

The paintings selected for the vignette of the dead Jesus come from artists who worked in the Baroque period (c. 1600 CE) when Christian subject matter was most common. In these paintings, light symbolized divine power and hope, while darkness (such as the dark wood from the beginning of Dante's *Commedia*) was a place of unknown and terrible things. Painting Jesus after his death presents a challenge for any artist because such a scene is centered upon an emaciated, wound-covered corpse. The artist is confronted by a scene that is, perhaps, the nadir of sorrow.

Contextually, light and dark are important as metaphors in the period between Christ's death and his resurrection. Matthew 27:45 describes how, for the hours of the day during the crucifixion, "there was darkness over all the land" (*King James Version*). In many paintings, much of the darkness is conveyed in the faces of those tending to Christ's body. Further, the divine light leaves Christ upon his death, yet his corpse is pale with the tinge of death. In the following paragraphs, two paintings of the lamentation are analyzed using the skeletonkey program.

Annibale Carracci's *The Lamentation* (see fig. 10) is as we would expect, appropriately mournful and dark. It presents a scene of undeniable, exhausted grief. The painting is filled with large, illuminated, looming figures. Yet, examining the histogram of the *adapthisteq* version reveals that the majority of the painting is in actuality nearly black. In the histogram, the highest bars are around a luminance of

50 to 100, which is very dark gray or black, confirming the overall dimness of this image. This fact, that the painting is mostly black even though there are large, illuminated figures, demonstrates how a quantitative analysis helps us consider the painting's overall brightness, to see more than what our eyes are trained to see.

Between the kneeling figures, boldly placed in the center of the painting, the artist has placed a shadowy recess like a deep abyss. The extent of this space's darkness is arresting. This centered shadow seems like a vast expanse over which the kneeling figures must gaze to see the lifeless body. The skeletonkey image shows it to be the darkest space, along with the upper-right corner. Chiaroscuro accounts for why this centered shadow seems especially black: it is keenly contrasted with the dead, wan body of Jesus.

The pixelated skeletonkey version also shows us that the black region at the top is counterbalanced by the lighter foreground near Christ and the brighter light on the mourner in the front who supports Christ's body. This achieves the effect of falling or descent. All of the light in the scene seems to rain down, wash over the bodies, and finally collect in a pool around Jesus in the form of the death shroud. The skeletonkey image makes it clear that the concentration of light areas is predominantly the triangle in the lower right. Our eyes follow the same path, looking from face to face and then down to the brightest spot, the cloth around the dead figure, the central element of the painting. Compositionally, the sense of falling is mirrored by the body of Christ and the figures holding him, all of whom seem to merge into a single entity. They have collapsed from bereavement. The two other women in the group also appear to have just fallen to their knees. The shadows envelop them on all sides, the right arms of both figures seeming to extend from a dark mist. The darkness oppressively and emphatically bears down on these women, reflected in their pained and tired faces. The adapthisteq image shows that each mouth is open in a cry, an exhale of futility and fatigue. If these same figures were not trapped in this shadowed world but were instead surrounded with gold leaf and light, the expressiveness of this scene might seem unduly exaggerated or satirical to our eyes. Instead, our perception of the painting is one of supreme disquiet as each figure laments, wrapped in a pall of hopeless gloom.

An earlier work concerning the lamentation, by Francesco Bassano the Younger (see fig. 11), takes the darkness around the dead Christ to an extreme. Bassano set his painting at night in a darkened place, lit by a sole candle. Demonstrating an awareness of composi-

tion, Bassano placed the focal point, the candle, just off center in the middle of his painting so that it would offset the darkness all around. The candle's presence, shining magnificently, could be seen to signify the hope still left even after Jesus' death, the hope of salvation and the return of the Holy Spirit. Though not as viscerally melancholy as Carracci's painting, Bassano's painting of the murky room instills the viewer with a sense of futility. The histogram shows that more than two-thirds of the painting is totally black. When this histogram is compared to that of Carracci's work, it is apparent that the distribution of lighter colors is more dramatic in the Bassano painting because the slope of its histogram is much steeper (a gentler slope in the Carracci may account for why that painting seems less dark than it truly is).

The adaptivesteq version of the Bassano painting reveals that the two forms in the back (on the right) truly float, materializing out of the darkness. The skeletonkey version makes it clear that they are no more than wisps, shrouded in darkness, away from the group, and facing each other. Their distant, shadowy presence could mean that they doubt or question faith. In contrast, the brightest figures from the skeletonkey image are the figures nearest the light, metaphorically nearest to God. Though dead, Jesus is the most illuminated, even unnaturally so, since part of his leg, pointed away from the candle (facing the viewer), is fully lit. The three women, likely the three Marys, are all relatively equally illuminated. In the skeletonkey version, the man wrapping the cloth around Christ and the figure just behind him are still discernible by small spots of bright white. Yet the aforementioned figures in the very back are just ghosts. They stand back from the corpse, apart from the candle and its light. Perhaps they, like the viewers of this painting, are merely observers, not connected to the rest by the candle's light, free to walk forward or to walk away. Compositionally, the presence of the figures provides a necessary balance to the work: they add some sense of depth to the blackness behind Mary. Figuratively, they seem to represent two whose faith has dimmed.

Together, these two pieces by Bassano and Carracci exemplify the way painters use chiaroscuro to create powerful mimetic works. On one hand, in Carracci's painting, the sense of a heavy doom is impressed on everyone, including the viewer, by the composition's use of prevalent, inescapable, and terrible shadows. On the other, the violent contrast between light and dark in Bassano's piece and its focus upon the brilliant flame, humble yet commanding, testify

to the symbolic might that lighting and shadows play in spiritual Christian works. It is perhaps most telling that in each painting the corpse is the least distressing element of the whole work: it is the light and dark together that affects us.

FUTURE DEVELOPMENT

AN EXTENSION OF this project would be to implement an online application of the algorithm, so that anyone could go to a website, upload any image of their choice, and investigate its light and dark structure at will. Such an endeavor would achieve the goal of disseminating the algorithm, encouraging more exploration and interaction between the viewers and the artworks.

CONCLUSION

BY COMBINING THE qualitative analytical methods of art history with the quantitative tools of image processing, “A Digital Skeleton Key to Art” has suggested a unique perspective on chiaroscuro in paintings, in the form of digital images, by using a computer program to reveal their structure, uncover hidden features, and elucidate the connotative meanings of those paintings.

Paintings contain more information, both physical and symbolic, than can be seen by a cursory or unaided glance. To examine paintings deeply is to befriend them. A different perspective, such as that afforded by the algorithm, allows a fresh way to understand the paintings. If we consider a painting the way a physician considers a patient, then we examine the skin (the painting’s color), the skeleton (the underlying drawing), and the functions of the organs (the interaction of the figurative elements). But if we consider a painting the way we consider another person, someone we truly wish to know, then we listen to it, observe it, and soon find ourselves drawn into its canvas, exploring its shadows and engaging its figures. Art history at its best is not staid and clinical but instead inspires active and personal connections between the artwork and the viewer.

Though they are abstractions, digital photographs of paintings contain the same hidden aspects that reside in the original. These aspects, such as a faded halo or the implication of a brilliant candle flame, are invisible to the eye unaided by either technology or history. By exploiting the abilities afforded by working with digital images,

the obscure facets and nuances of shadow in paintings can emerge. Rather than regarding them as antithetical to the nature of art historical investigation, we should use digital image processing tools to augment the analysis of art. These tools are already in our hands; the responsibility is ours to make the most of them.



FIG. 1 Source: Unknown Artist. *Hades Abducting Persephone*. c. 300 BCE. Vergina, Macedonia. *Wikimedia Commons*. Web. 26 Nov. 2011.



FIG. 2 Source: Michelangelo da Caravaggio. *Supper at Emmaus*. 1601. National Gallery, London. *Wikimedia Commons*. Web. 26 Nov. 2011.



FIG. 3 Source: Rembrandt van Rijn. *Self-portrait*. 1628. Rijksmuseum, Amsterdam. *Wikimedia Commons*. Web. 26 Nov. 2011.



FIG. 4 Source: Jacques-Louis David. *The Lictors Bring to Brutus the Bodies of His Sons*. 1789. Louvre Museum, Paris. *Wikimedia Commons*. Web. 26 Nov. 2011.

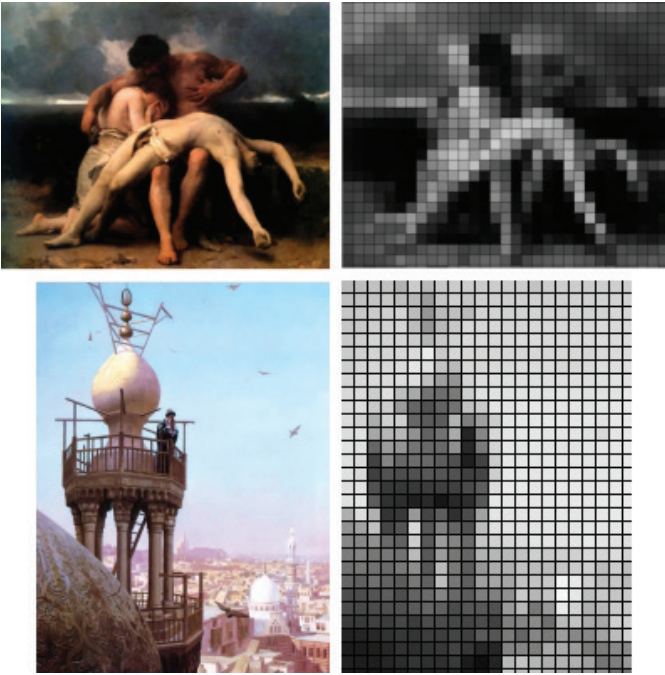


FIG. 5 Original and Prototype Images. Source (top): William-Adolphe Bouguereau. *The First Mourning*. 1888. National Museum of Fine Arts, Buenos Aires. *Wikimedia Commons*. Web. 26 Nov. 2011. Source (bottom): Jean-Léon Gérôme. *A Muezzin Calling from the Top of a Minaret the Faithful to Prayer*. 1879. Private Collection. *Wikimedia Commons*. Web. 26 Nov. 2011.

1. Load image, convert to grayscale if needed
2. Collect data: get the dimensions, other pixel data
3. Draw grid of cells over image
4. Abstract the cells so each is a single color
5. Collect data: e.g. histogram of total lightness
6. Repeat steps 3 - 5 as needed
7. Determine lightest and darkest region
8. Explain symbolism through analysis of the image and data
9. Compare to other images, tie it into the 'vignette' (the theme)

FIG. 6 Procedure to Abstract Images

```

% This is the skeletonkey function
% Author: John Winder
% Date: 2010-2011
% Inputs:
% filename, the name of the color image to be used (a string);
% step, the number of pixels in the length of a single cell
% in the generated image (an integer, minimum is 1)
% Outputs:
% out, a matrix containing the modified image data
% Example:
% M = skeletonkey('mona_lisa.jpg', 1);
% would store image data identical to mona_lisa.jpg in matrix M
% (because the step is 1, there is no abstraction)
% N = skeletonkey('mona_lisa.jpg', 25);
% produces the image data for a pixelated mona_lisa.jpg where the
% cells have lengths of 25 (storing image data in N).
function out = skeletonkey(filename, step)
A = imread(filename);
A = rgb2gray(A); % The assumption is that color images are used.
B = adapthisteq(A); % This is the highly contrasted version
[fileM, fileN] = size(A); % Storing the dimensions of the image data
for x = 1:step:fileM
    xLower = x;
    xUpper = (x + step) - 1;
    if xUpper >= fileM
        xUpper = fileM;
    end
    for y = 1:step:fileN
        yLower = y;
        yUpper = (y + step) - 1;
        if yUpper >= fileN
            yUpper = fileN;
        end
        tSection = A(xLower:xUpper, yLower:yUpper);
        tTotal = sum(sum(tSection));
        [tM, tN] = size(tSection);
        tAverage = tTotal / (tM * tN);
        tSection = tAverage * ones(size(tSection));
        A(xLower:xUpper, yLower:yUpper) = tSection;
    end
end
% display the skeletonkey version and its histogram
% then display the enhanced version and its histogram
figure; imshow(A), figure; imhist(A), figure; imshow(B), figure; imhist(B);
out = A;
end

```

FIG. 7 The file for skeletonkey.m. This program generates the unlocked and enhanced images as well as the corresponding histograms.



FIG. 8 At the right is the adapthisteq version. Source: Franciabigio. *Virgin and Child with Job and John the Baptist*. 1516. Museo di San Salvi, Florence. *Wikimedia Commons*. Web. 26 Nov. 2011.

FIG. 9 At the right is the skeletonkey version. Source: Michelangelo da Caravaggio. *Ecce Homo* (detail). 1605. Palazzo Rosso, Genoa. Wikimedia Commons. Web. 26 Nov. 2011.

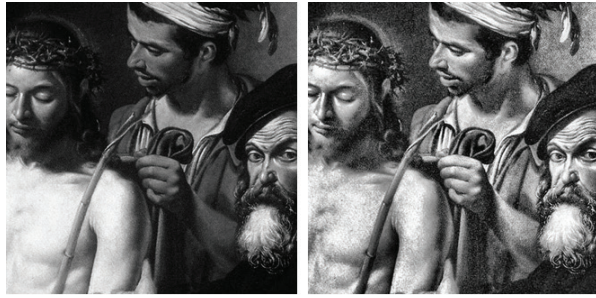


FIG. 10 At the lower left is the adapthisteq version. At the lower right is the skeletonkey version. The bars of the histogram show the pixel count for each value along the spectrum. Source: Annibale Carracci. *The Lamentation*. 1606. National Gallery, London. Wikimedia Commons. Web. 26 Nov. 2011.

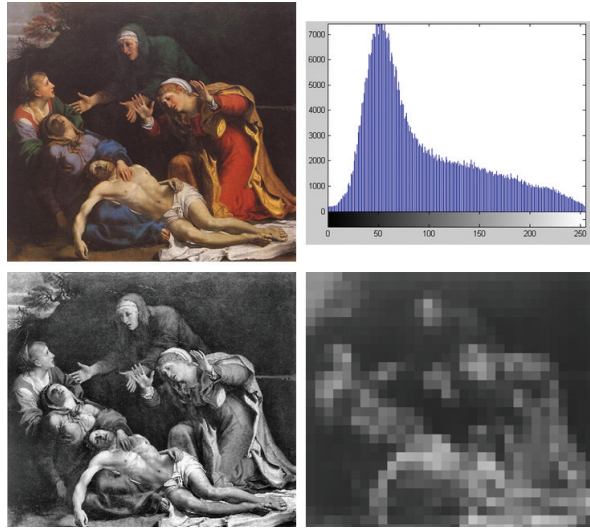
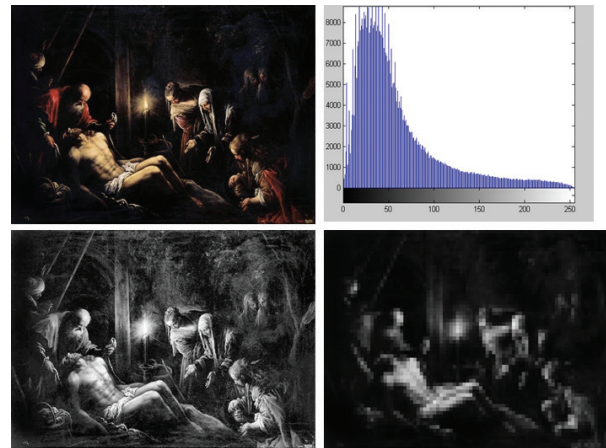


FIG. 11 At the lower left is the adapthisteq version. At the lower right is the skeletonkey version. The bars of the histogram show the pixel count for each value along the spectrum. Source: Francesco Bassano the Younger. *The Lamentation*. c. 1580. Private Collection. Wikimedia Commons. Web. 26 Nov. 2011.



WORKS CITED



- Baxandall**, Michael. *Shadows and Enlightenment*. New Haven: Yale University Press, 1995. Print.
- Bell**, Janis C. "Light and Color in Caravaggio's 'Supper at Emmaus'." *Artibus et Historiae* 16.31 (1995): 141-142. JSTOR. Web. 24 Mar 2011. <<http://www.jstor.org/stable/1483502>>.
- Cavanagh**, P. "Reconstructing the Third Dimension: Interactions between Color, Texture, Motion, Binocular Disparity, and Shape." *Computer Vision, Graphics, and Image Processing* 36.2-3 (1986): 395. Web. 2 Apr. 2011. <<http://visionlab.harvard.edu/Members/Patrick/PDF.files/2003%20pdfs/recon.pdf>>.
- "**Chiaroscuro**." *The Grove Encyclopedia of Materials and Techniques in Art*. Oxford: Oxford UP, 2008. Print.
- Culicover**, Peter W. "Squinting at Dali's Lincoln: On how to think about language." Proceedings of the 41st Annual Meeting of the Chicago Linguistics Society. Chicago Linguistics Society, Chicago. (2008). Web. 24 Mar. 2011. <<http://www.ling.ohio-state.edu/~culicove/Publications/Lincoln.pdf>>.
- Fairchild**, Mark D. *Color Appearance Models*. Reading, Mass.: Addison-Wesley, 1998. Web. 24 Mar. 2011. <http://books.google.com/books?id=8_TxzK2B-5MC>.
- Gombrich**, Ernst H. *Shadows: the Depiction of Cast Shadows in Western Art*. London: National Gallery Publications; 1995. Print.
- Gonzalez**, Rafael C., Richard E. Woods, and Steven L. Eddins. *Digital Image Processing Using MATLAB*. Upper Saddle River, NJ: Pearson, 2004. Print.
- The Holy Bible: King James Version*. Nashville: The Gideons International, 1978. Print.
- Keuls**, Eva. "Skiagraphia Once Again." *American Journal of Archaeology* 79.1 (1975): 1-16. JSTOR. Web. 24 Mar. 2011 <<http://www.jstor.org/stable/503335>>.
- "**Painting**, Western." *Encyclopædia Britannica*. Standard Edition. 2011. Web. 24 Mar. 2011.
- Stoichita**, Victor I. *A Short History of the Shadow*. London: Reaktion Books, 1997. Print.

MODELING A CELLULAR RESPONSE TO A GRADIENT

Mathematics and Molecular Biology Inform a
Mechanistic Understanding

**xuan ge,
david stonko**

Xuan Ge is a biology major with a minor in psychology who will be graduating in the spring of 2012. She is a member of the Honors College as well as a recipient of the UMBC Premier Award. She plans to continue on to medical school after UMBC. David Stonko is a junior math and biology double major who hopes to continue his education at medical school after graduation from UMBC in spring 2013. David is a member of the Undergraduate Training Program in Biology and Mathematics (UBM) and various other student organizations. Xuan and David would like to thank their research mentors, Dr. Michelle Starz-Gaiano and Dr. Bradford Peercy, for their continued guidance, patience, and encouragement throughout the course of this project. They would also like to thank everyone in the UBM program and their friends, family, and teachers who have supported them.



We began our research after we were selected to participate in the Undergraduate Training Program in Biology and Mathematics (UBM) supported by the National Science Foundation. This program was created to promote the integration of mathematics in biology and to foster an appreciation of interdisciplinary research. Our project is concerned with proposing and testing a new biological hypothesis while building our mathematical model. The entire research process has been an eye opening experience and we believe everyone in this program, as well as other undergraduates in research, can agree that it has helped to better shape our understanding and goals here at UMBC.

INTRODUCTION

CELL MIGRATION IS prevalent in normal development of all animals, and pathological conditions like birth defects or metastatic cancers can arise when this process goes awry. Understanding the phenomenon of cells undergoing the transition from a stationary state to a migratory stage is of broad interest. Such a complex problem can be more easily studied in a simple organism, such as the fruit fly *Drosophila melanogaster* (Naora and Montell, 2005). *Drosophila* is one of the most studied organisms in biological research. Its short generation time, high fecundity, visible congenital markers, and well-characterized genome are just a few of the many traits that make it an ideal organism for genetic studies. In addition, about eighty percent of disease genes in humans encode proteins that are conserved in flies (Reiter et al., 2001). Thus, because the underlying molecular signaling mechanisms are well conserved, insight into cell migration gleaned from *Drosophila* will likely be broadly applicable.

We are interested in how a set of cells in flies, called the border cells, becomes motile. The border cells arise as part of the follicular epithelium during oogenesis, but change character from epithelial to migratory and travel between other cells to the oocyte, where they are required for embryo development (Fig 1A) (Jang et al., 2007; Montell, 2008). A single egg chamber of *Drosophila* consists of an oocyte accompanied by fifteen nurse cells, and together, they are surrounded by an epithelium of follicle cells. Early in the development of the egg chamber, a pair of cells on each end specializes into the polar cells. In mid-oogenesis, the polar cells at the anterior end release molecular signals that induce six to eight neighboring follicle cells to become border cells (Fig 1A-E). A few hours later, the border cells will withdraw from the epithelium and escort the polar cells a distance of about 20 cell diameters to the edge of the oocyte by stage ten (Fig 1D-E). The follicle cells' transformation into mobile border cells requires changes in gene expression that allow their detachment from neighboring cells and production of cytoskeletal changes that promote migration. Because a similar epithelial to mesenchymal transition underlies tumor metastasis, the analysis of *Drosophila* border cells not only allows us to expand our current knowledge of this particular molecular signaling system, but may also yield insight into mechanisms of human disease progression.

A CASCADE OF MOLECULAR SIGNALS ACTIVATES CELL MOTILITY

THE JANUS KINASE (JAK)/Signal Transducer and Activator of Transcription (STAT) signaling pathway has been shown by previous studies to be intimately involved in the border cells' acquisition of mobility (Arbouzova and Zeidler, 2006; Beccari et al., 2002; Ghigliione et al., 2002; Silver and Montell, 2001; Xi et al., 2003). The anterior polar cells of the *Drosophila* egg chamber secrete the cytokine Unpaired, which acts as the ligand for the transmembrane Domeless receptor in neighboring follicle cells. The binding of Unpaired to Domeless activates JAK which leads to its phosphorylation. Once phosphorylated, the activated JAK/receptor complex recruits and phosphorylates STAT. The phosphorylated STAT is then able to dimerize, move to the nucleus, and act as a transcription factor to turn on specific target genes. As indicated by Fig 1, the released signal, Unpaired, diffuses outwards from the polar cells and forms a gradient across the neighboring epithelium. This ensures that the cells in closer proximity to the polar cells experience a higher level of the ligand and turn on a higher level of STAT activity, which promotes their transformation into mobile border cells. The gradient of STAT activity was directly observed using fluorescent proteins in the egg chambers and standard optical sectioning (Fig 1F), and is essential for the specification of the right number of motile cells.

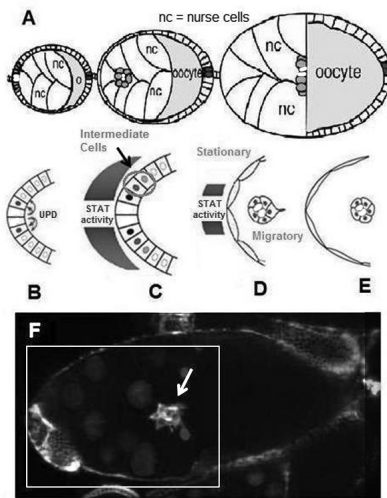


FIG. 1 Egg development and the migration of border cells. Panel A is a cartoon showing the maturation process of a *Drosophila* egg chamber and the relative movement of the border cells between the nurse cells (nc). Panels B through E are a magnification of the anterior quarter of an egg chamber during normal development, in which the signaling molecule Unpaired is secreted from the polar cells (B) and induces a gradient of STAT activity across the anterior epithelium at stage 8 (C). By late stage 9, STAT activity is limited to the follicle cells very close to the polar cells, which assume the identity of border cells, become motile (D), and migrate towards the oocyte (E). Panel F is a Fluorescent Protein-STAT reporter (from (Baeg et al., 2005)) staining of a wild type egg chamber, which demonstrates that STAT transcriptional activity is highest in the migrating border cell cluster (arrow). Panel E corresponds to the boxed region of panel F.

Studies by several labs (Beccari et al., 2002; Ghiglione et al., 2002; Silver and Montell, 2001), showed that the loss of STAT signaling in the follicle epithelium produced no migratory cells, and that a higher STAT signaling in the follicle cells induced more cells to migrate. In 2005, Silver et al. showed that the proper level of STAT signaling was required during the cell movements as well as in the initial specification (Silver et al., 2005). These findings suggest that precise regulation of JAK/STAT signaling is necessary to generate the migratory behavior in the correct number border cells and that it can induce motile behavior in other follicle cells.

Interestingly, the normal decision of border cells to migrate, while other cells in the egg chamber do not, cannot be controlled by Unpaired and STAT alone. In fact, there are other signaling molecules, discovered through genetic analysis, such as SLBO and Apontic, involved in the JAK/STAT signaling pathway that also play integral roles in defining the trajectory of border cell development (Starz-Gaiano et al., 2008).

TRIGGER MOLECULE FOR MIGRATION: SLBO

SLOW BORDER CELLS, or SLBO, is an important downstream transcriptional target of activated STAT and promotes the migratory behavior. Montell, Rorth, and colleagues (1992) demonstrated that insufficient levels of SLBO suppress border cell migration and therefore prevent the formation of viable eggs (Montell et al., 1992). The *slbo* gene contains several binding sites of varying affinity for STAT, which causes SLBO expression to follow the same graded pattern as JAK/STAT in normal development, with the highest concentrations in cells adjacent to the polar cells early (as indicated by Green Fluorescent Protein (GFP) reporters (Starz-Gaiano et al., 2008)). A few hours later, however, STAT and SLBO are expressed in a step-wise manner with high levels in cells that become migratory and none in other follicle cells. This observation suggests that high levels of SLBO are positively correlated with high expression of STAT, promoting the migratory pathway, and that this signal must be shut off in non-motile cells.

INHIBITOR MOLECULE FOR MIGRATION: APONTIC

APONTIC IS A transcription factor that opposes the function of JAK/STAT and SLBO, and functions to inhibit migration (Fig A: Starz-Gaiano et al., 2008). It is also a downstream target of STAT transcriptional activity, but is additionally turned on by another transcription factor; thus, it is more evenly expressed than SLBO within the domain of the STAT activity gradient. The apontic mutant phenotype in egg chambers is characterized by the migration of additional cells trailing behind the main mobile group, as demonstrated in Fig 2B. This loss-of-function phenotype closely resembles the overstimulation of JAK/STAT, which strongly suggests that normally, Apontic has antagonizing effects on JAK/STAT activity. Prior work shows that Apontic acts as a feedback inhibitor on the JAK/STAT pathway, and that this process is mediated by Apontic's activation of microRNAs that reduce STAT protein (Yoon et al., 2011).

It is clear that Apontic and SLBO inhibit the expression of one another (Starz-Gaiano et al., 2008). Unlike the wild type, apontic mutants maintained high levels of SLBO in follicle cells farther away from the polar cells, and therefore induced the migration of additional cells. In wild-type egg chambers, all eight migrating border cells express SLBO, but when Apontic was overexpressed SLBO protein was undetectable. Additional experiments showed that APT can directly repress *slbo* transcription. SLBO inhibits the expression of Apontic also, but to a lesser degree, since SLBO overexpression only decreases the levels of Apontic instead of blocking it completely, and expression from the *apt* locus is unchanged. The mechanism for SLBO inhibition of APT is unknown.

BUILDING A MATHEMATICAL MODEL OF A MOLECULAR NETWORK THAT GOVERNS TWO CELL FATES

A SIMPLE CIRCUIT demonstrates the positive and negative regulatory relationships among all the factors involved in the JAK/STAT signaling system as well as the fates they promote, based on genetic and biochemical data (Fig 2A). However, this representation does not explain well how some cells with STAT activity promoted sig-

naling through APT to remain stationary while others reinforced SLBO activation and became motile.

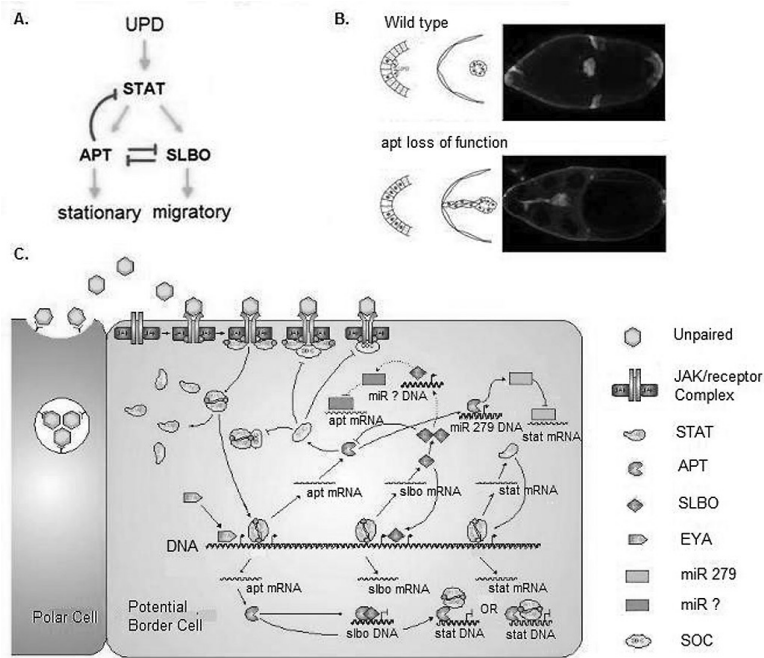


FIG. 2 Genetic components that regulate migratory cell fate. Panel A shows the initially proposed biological circuit demonstrating the stimulatory and inhibitory relationships among STAT, SLBO, and Apontic, along with the outcomes they induce (Starz-Gaiano et al., 2008). Apontic inhibits, but is also stimulated by STAT. The high expression of Apontic contributes to the stationary phenotype and inhibits SLBO expression as well. STAT upregulates SLBO, which in turn promotes the migratory behavior. High SLBO represses Apontic expression. The top figures of Panel B illustrate the migration of wild type border cells out of the epithelium. The bottom figures of Panel B show a mutant phenotype, in which a loss of function mutation in Apontic induces the migration of additional cells. The top figures demonstrate the migration of wild type border cells, and they can be found by the edge of oocyte by early stage 10. In comparison, the bottom figures of Panel B show the trailing of additional cells behind the migrating cluster due to the loss of function of Apontic (Starz-Gaiano et al., 2009; Starz-Gaiano et al., 2008). Panel C shows the more comprehensive biological schematic we build upon the simple circuit in Panel A, and was informed by multiple research articles. This schematic attempts to address all the molecular interactions among the factors involved as well as the stimulatory and inhibitory interchanges between them.

How do cells set a threshold to determine the level of signal that is sufficient to promote movement? We observed that initially STAT-positive cells could acquire either motile or non-motile states, and this led us to consider which known components were essential to act as a switch. The simple biological circuit previously described (Starz-

Gaiano et al., 2009; Starz-Gaiano et al., 2008) did not describe all the possible molecular interactions among the components. Similarly, even though the previous mathematical model qualitatively captures the behaviors of the parameters involved, it is a heuristic model that cannot be used to understand or predict the molecular interactions of the system and direct future biological study. To circumvent these problems, we first developed a more comprehensive biological schematic (Fig 2C), and then, developed a mechanistic mathematical model that builds on the elementary reactions, which can be used to determine which components may be more important for understanding the cell fate switching behavior.

UNDERSTANDING THE THRESHOLD LEVEL OF MOLECULAR SIGNALING ESSENTIAL FOR MOTILITY

ALL FOLLICLE CELLS that receive extracellular Unpaired activate STAT, but cells that are closest to the polar cells (the source of UPD) have high levels of STAT activity and cells that are far away have low STAT activity. It is the cells at an intermediate distance from the polar cells that are of greater interest to us because these are the cells in which cell fate determination becomes crucial. Mutant analysis has shown that disruption of STAT activity in these cells results in poorly specified, poorly motile cells, as is shown in Fig 2B. At this distance the cells turn on an intermediate level of STAT activity, which could either fall above or below the threshold for transition, which is a precise level of STAT activation. Cells that do not surpass the threshold will eventually turn off all STAT activity and remain stationary, whereas cells that do meet the requirement will amplify the signal to become migratory. Because these intermediate follicle cells have the potential to achieve two different fates, we can treat their transition as a bistable system.

The concept of a threshold and bistability allow us to develop a model that consists of two stable steady states for JAK/STAT activity. For our purposes, the rate at which Unpaired is degraded is kept at a constant due to the presumed similarity in metabolic rates of the cells under investigation. The cells in closer proximity to the polar cells are more likely to exceed the threshold of STAT activation than cells farther away, and therefore they transform into mobile border cells, which travel with the polar cells to the periphery of the egg

chamber. There are two stable steady states in JAK/STAT. The first steady state, due to the negligible concentration of Unpaired at time zero, represents an absence of JAK/STAT activity. The second steady state, due to the high concentration of Unpaired at some long time after its initial release, represents the amount of JAK/STAT activity to allow the migratory transition. This model is consistent with our understanding that any subthreshold signaling, as experienced by follicle cells far away from the polar cells, would cause the system to fall back to the first steady state and compel the follicle cells to remain epithelial. Any suprathreshold signaling, as experienced by cells immediately adjacent to the polar cells, would push the system to the second steady state, enabling enough JAK/STAT activity to initiate the migratory pathway and generate mobile border cells that carry the polar cells to the boundary of the oocyte. The details of this dynamic will be explained further in the sections below.

INTERMEDIATE FOLLICLE CELLS ATTENUATE STAT SIGNALING IN THE ABSENCE OF MIGRATION

IN A WILD type *Drosophila* egg chamber, STAT activity initially forms a gradient across the anterior epithelium (Starz-Gaiano et al., 2008; Xi et al., 2003). This graded pattern is mirrored closely by SLBO expression since STAT acts as a transcriptional activator to increase SLBO expression. Depending on whether or not cells reach the necessary high level of STAT activity, they may or may not prepare to make the transition into motile border cells. As described above, intermediate follicle cells (which do not migrate) initially have STAT activity, but then shut it off in an APT- dependent manner as the border cells move away.

Possibly, the down-regulation of STAT activity could merely be a result of the activator, UPD secreted from the polar cells, being removed as the border cells migrate. To test this idea, we examined carefully-staged egg chambers just before migration takes place, and assayed STAT activity by expression of SLBO. We found that only the rounded-up border cells, and not the intermediate cells, expressed SLBO (Fig 3), even when they were a similar distance from the polar cells. This supports the idea that the gradient of biological activator forms before migration takes place, that migration is not required for the establishment of two distinct cell fates (establishment of a

step-wise STAT activation), and that the intermediate cells have an intrinsic mechanism for down-regulating sub-threshold levels of STAT activity.

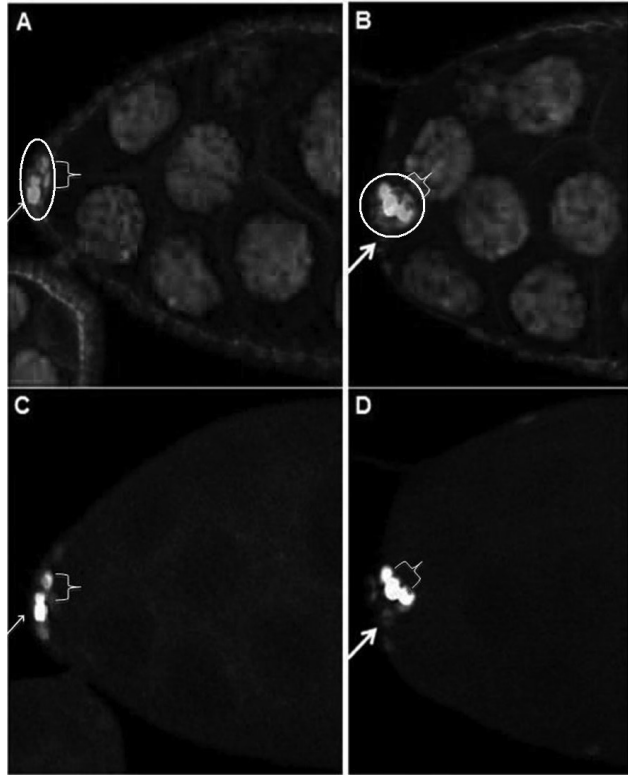


FIG. 3 Intermediate cells downregulate STAT prior to the activating signal being removed. Grayscale images of fluorescent antibodies that label the border cells. All cells receiving STAT signal express some level of SLBO protein (shown alone in bottom panels). DAPI staining indicating nuclear DNA is shown for all cells, and the cell-surface marker, Armadillo, highlights the border cells (circled) in top panels. Panel A and C show a wild type egg chamber at late stage 8. Panel B and D show an egg chamber at early stage 9, in which the border cells have already clustered at the anterior end, ready to migrate. These cells maintain a high level of STAT activity, as demonstrated by the high SLBO expression. The cell indicated by the arrow represents a cell that once turned on STAT, and remains physically close to the activating signal from the polar cells (bracket), but is no longer activated due to its inability to reach the threshold. attempts to address all the molecular interactions among the factors involved as well as the stimulatory and inhibitory interchanges between them.

We were interested in understanding potential cell-intrinsic mechanisms that could generate two major differences in cell fates in response to very small differences in STAT activity levels. Our current model of the JAK/STAT signaling pathway is based on the law

of mass action, which attempts to address not only the positive and negative relationships between regulators, but also the molecular interactions that exist among these players. By invoking the biological function of the players into the model, we can gain a more mechanistic insight of the system, and establish a more comprehensive understanding of the migratory transition of *Drosophila* border cells. Thus, we explored the possibility of achieving a bistable system with a minimal number of components using our mathematical model. We demonstrated that bistability can be achieved by the cross-repressional interactions between APT and SLBO.

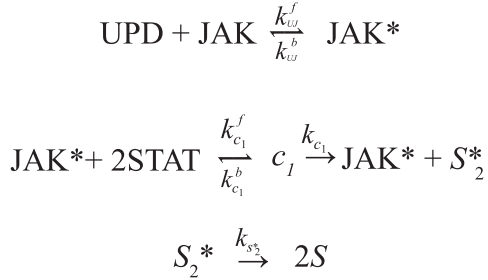
MOTIVATION AND CONSTRUCTION OF A MATHEMATICAL MODEL

PREVIOUS WORK HAS been done heuristically modeling the system described in Fig 2A (Starz-Gaiano et al., 2009; Starz-Gaiano et al., 2008; Yoon et al., 2011). However, while the functional forms of those equations capture the positive or negative interactions between species, those forms are generic and not necessarily the most appropriate forms for the system. Our goal is to utilize the elementary reactions of this system leading to a more detailed, mechanistic mathematical model that captures the underlying architecture of the system.

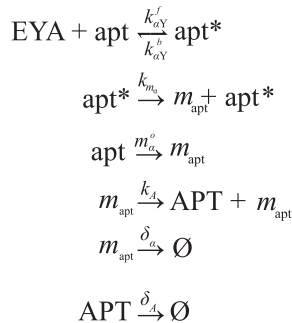
We develop a set of reaction equations that captures the underlying mechanisms controlling the JAK/STAT pathway. Using this set of reactions, we create a system of differential equations that tracks the production and activation of the molecules in the pathway. We relate this new mechanistic mathematical model to the biological system to identify where the terms originate and how they impact the behavior of the model, where previously it was not possible to relate each term in the equations back to the biological model. The approach results in new insight into the mechanisms of cell motility and helps to inform new testable hypotheses.

The Law of Mass Action is a mathematical representation of chemical reactions that is useful in predicting and explaining the behavior of interacting molecules in a solution (Keener and Sneyd, 1998). Using the reaction scheme in Fig 2C we developed, using the Law of Mass Action, a system of differential equations that track each molecule through time. We systematically analyze each component of the system and write down its reaction equation.

We first consider the activation of the pathway by the ligand, UPD. UPD binds JAK to form activated JAK (JAK*) at a rate k_{UJ}^f . The reverse reaction occurs at a rate k_{UJ}^b . This activated JAK then binds two STAT monomers to form a complex, c_1 . An activated STAT dimer (referred to as S_2^*) dissociates from this complex, leaving activated JAK, at a rate k_{c_1} . The dedimerization of activated STAT dimer back into STAT monomers occurs at a rate $k_{S_2^*}$. We have written the following set of reaction equations:

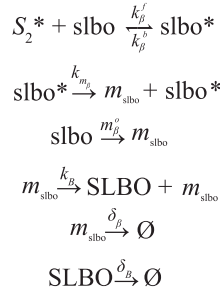


This dimerized STAT enters the nucleus and activates transcription of *apt* upon binding its enhancer. Here, *apt*'s sensitivity to activation by STAT is described by k_a , which is equal to the ratio of transition rates of activation and deactivation, k_a^f/k_a^b . Transcription of *apt* can also be activated by the protein Eyes absent (EYA). The gene is transcribed into mRNA at a rate k_{m_a} . The *apt* gene may also be activated by other, STAT independent factors, producing a basal level of *apt* mRNA at a low, constant rate, m_a^o . The mRNA from all sources is then translated into protein at a rate k_A . Over time the mRNA and protein degrade at rates δ_a and δ_A , respectively. These interactions with APT are captured in the reactions:

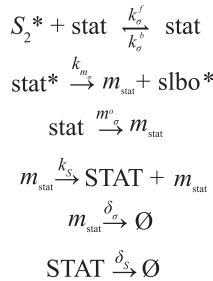


Similarly, transcription of *slbo* is activated by STAT binding and the rate of activation also has a certain sensitivity to STAT,

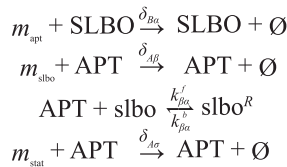
k_β . The activated gene is transcribed into mRNA, m_{slbo} which is then translated into SLBO protein. The mRNA and gene product both degrade in the cell at rates δ_β and δ_B , respectively. These interactions with SLBO are captured in the reactions:



STAT protein (monomer) is also produced auto catalytically. Activated, dimerized STAT binds its own enhancer and initiates transcription. The *stat* gene is also assumed to be activated at low levels by other transcription factors producing STAT mRNA (m_{stat}), at a rate m_{stat}^o , and which is assumed to degrade in the cell at a rate δ_σ . The following reactions describe these interactions:



We also model the cross inhibition of SLBO and APT and the feedback inhibition of APT on STAT using the same method. SLBO represses APT post-transcriptionally while APT antagonizes SLBO by repressing its transcription. APT has also been shown to turn on transcription of a microRNA intermediate that disrupts the transcription of the *stat* gene. These interactions can be written as:



This complete reaction scheme was rewritten as a system of differential equations, which describes the production, activation, inactivation, and degradation of each chemical in time based on the reaction equations. Fig 4 shows the full system of differential equations (1) - (15). In each equation, the production, activation, or inactivation is modeled with A and B denoting APT and SLBO, with m denoting the mRNA of the gene defined in the subscript (e.g., mRNA of *apt* is written as m_α , where the subscript α implies *apt*). Furthermore, α, β and σ represent the fraction of their respective gene that is transcriptionally inactive. We assume conservation so that the proportion of each gene that is inactive, repressed, and active totals one.

$$\begin{aligned}
 (1) \quad & \frac{dJ^*}{dt} = k_{U,J}^f UJ - k_{U,J}^b J^* - k_{c_1}^f J^* S^2 + k_{c_1}^b c_1 + k_{c_1} c_1 \\
 (2) \quad & \frac{dJ}{dt} = -k_{U,J}^f UJ + k_{U,J}^b J^* \\
 (3) \quad & \frac{dS}{dt} = -2k_{c_1}^f J^* S^2 + 2k_{c_1}^b c_1 + 2k_{S_2} S_2^* + k_S m_\sigma - \delta_S S \\
 (4) \quad & \frac{dc_1}{dt} = k_{c_1}^f J^* S^2 - k_{c_1}^b c_1 - k_{c_1} c_1 \\
 (5) \quad & \frac{dc_3}{dt} = k_{S_2 A}^f S_2^* A - k_{S_2 A}^b c_2 \\
 (6) \quad & \frac{dS_2^*}{dt} = k_{c_1} c_1 - k_{S_2} S_2^* - k_{S_2 A}^f S_2^* A + k_{S_2 A}^b c_2 \\
 (7) \quad & \frac{dA}{dt} = k_A m_\alpha - \delta_A A - k_{S_2 A}^f S_2^* A + k_{S_2 A}^b c_2 \\
 (8) \quad & \frac{dB}{dt} = k_B m_\beta - \delta_B B \\
 (9) \quad & \frac{dm_\alpha}{dt} = k_{m_\alpha} (1 - \alpha) - \delta_{m_\alpha} m_\alpha + m_\alpha^o - \delta_{B\alpha} B^2 m_\alpha \\
 (10) \quad & \frac{dm_\beta}{dt} = k_{m_\beta} (1 - \beta - \beta^R) - \delta_{m_\beta} m_\beta + m_\beta^o - \delta_{A\beta} A m_\beta \\
 (11) \quad & \frac{dm_\sigma}{dt} = k_{m_\sigma} (1 - \sigma) - \delta_{m_\sigma} m_\sigma + m_\sigma^o - \delta_{A\sigma} A m_\sigma \\
 (12) \quad & \frac{d\alpha}{dt} = -k_{\alpha}^f S_2^* \alpha + k_{\alpha}^b (1 - \alpha) - k_{\alpha Y}^f E \alpha + k_{\alpha Y}^b (1 - \alpha) \\
 (13) \quad & \frac{d\beta}{dt} = -k_{\beta}^f S_2^* \beta + k_{\beta}^b (1 - \beta - \beta^R) + k_{\beta A}^b \beta^R - k_{\beta A}^f A \beta \\
 (14) \quad & \frac{d\beta^R}{dt} = k_{\beta A}^f A \beta - k_{\beta A}^b \beta^R \\
 (15) \quad & \frac{d\sigma}{dt} = -k_{\sigma}^f S_2^* \sigma + k_{\sigma}^b (1 - \sigma)
 \end{aligned}$$

FIG. 4 Full system of differential equations used to model the JAK/STAT pathway. Each equation tracks the production, activation, inactivation and degradation of one component of the pathway shown in Fig 2C. Throughout the system a superscript of an asterisk implies activation and a superscript of an 'R' denotes repression.

ESTABLISHMENT AND ANALYSIS OF BISTABILITY IN THE MATHEMATICAL MODEL

THE PREVIOUS MATHEMATICAL model (Starz-Gaiano et al., 2008) was constructed to exhibit certain observed features of the biological system, such as bistability. Because our model was built from the elementary interactions, we began our mathematical analysis of our model by determining if it exhibited bistability for any family of parameters. In addition to determining if the system was bistable, we also wanted to isolate which components of the system were sufficient for bistability. We suggest that the simple cross repression system of APT and SLBO (depicted in Fig 5) is the key factor in causing bistability. To determine if this minimal system was responsible for bistability we simplified our equations to only include those components involved in the cross repression.

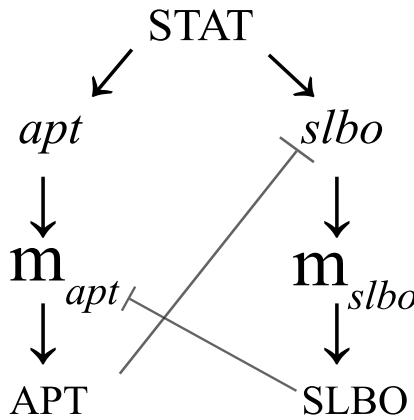


FIG. 5 Genetic circuit explaining the minimal cross repression system used to investigate bistability. SLBO represses *apt* at the gene level and APT represses SLBO at the mRNA level. Arrows indicate activation and lines indicate repression.

We isolate the equations that show STAT activating *apt* and *slbo*, their transcription into mRNA, and translation into protein. We also incorporate APT's inhibition of *slbo* at the gene/mRNA level and SLBO inhibition of APT post-transcriptionally (Starz-Gaiano et al, 2008). This gives the following system of differential equations to describe the simplified system in Fig 6, (16) - (22).

$$(16) \quad \frac{dA}{dt} = k_A m_\alpha - \delta_A A$$

$$(17) \quad \frac{dB}{dt} = k_\beta m_\beta - \delta_B B$$

$$(18) \quad \frac{dm_\alpha}{dt} = k_{m_\alpha} (1 - \alpha) - \delta_{m_\alpha} m_\alpha + m_\alpha^o - \delta_{B\alpha} B^2 m_\alpha$$

$$(19) \quad \frac{dm_\beta}{dt} = k_{m_\beta} (1 - \beta - \beta^R) - \delta_{m_\beta} m_\beta + m_\beta^o - \delta_{A\beta} A m_\beta$$

$$(20) \quad \frac{d\alpha}{dt} = k_\alpha^f S_2^* \alpha + k_\alpha^b (1 - \alpha)$$

$$(21) \quad \frac{d\beta}{dt} = k_\beta^f S_2^* \beta + k_\beta^b (1 - \beta - \beta^R)$$

$$(22) \quad \frac{d\beta^R}{dt} = k_{\beta^R}^f A \beta - k_{\beta^R}^b \beta^R$$

FIG. 6 System of differential equations that captures the simple cross repression system. Note that cooperativity of SLBO to repress APT was found to be essential for bistability and is included in the last term of equation (4).

In order to determine the parameters that produce bistability, we placed this system of differential equations into a steady state (i.e., all of the derivatives are set to zero). We then solved the system of equations for A and B . This gives us the following set of equations.

$$(23) \quad A = \frac{k_A}{\delta_A} \frac{k_{m_\alpha}}{\delta_{m_\alpha}} \left[\frac{\frac{S_2^*}{S_2^* + k_\alpha} + \frac{m_\alpha^o}{k_{m_\alpha}}}{1 + \frac{\delta_{B\alpha}}{\delta_{m_\alpha}} B^2} \right]$$

$$(24) \quad B = \frac{k_B}{\delta_B} \frac{k_{m_\beta}}{\delta_{m_\beta}} \left[\frac{\frac{S_2^* + A/k_{\beta R} (1/k_\beta^f - 1)}{S_2^* + k_\beta + A/k_{\beta R} (1/k_\beta^f - 1)} + \frac{m_{\beta^o}}{k_{m_\beta}}}{1 + \frac{\delta_{A\beta}}{\delta_{m_\beta}} A} \right]$$

FIG. 7 Steady-state equations

Then we substituted the first equation above into the second equation to generate one equation for B in terms of B . When the left side of this equation is equivalent to the right side the system is in a state of equilibrium. Our goal was to find biologically reasonable param-

eters that achieve three mathematical equilibria, two stable equilibria separated by one unstable equilibrium, indicating bistability. We also sought a system in which decreasing STAT would put the system in only one low steady state, while increasing STAT would lead to one elevated steady state.

Our initial equations (not shown) were unable to produce these results. However, a review of our model and further testing revealed that if we included cooperativity of SLBO to repress APT, then the system exhibits the ability to be bistable where without such a change, the system could never be bistable for any set of parameters. The parameters that we used for the basal bistable system are provided in Table 1.

Initial Parameters		
Parameter	Symbol	Value
Rate of APT Translation	k_A	0.5
Rate of transcription of apt	k_{m_a}	0.9
Rate of STAT independent production of APT mRNA	m_a^o	0.52
Rate of degradation of APT	δ_A	1
Rate of degradation if mRNA of APT	δ_{m_a}	0.05
Degradation rate of mRNA of APT due to miRNAs	$\delta_{A\sigma}$	0.1
Binding rate of STAT to apt and dissociation rate	k_a^f	1
	k_a^b	9
Rate of SLBO translation	k_B	2
Rate of transcription of slbo	k_{m_β}	0.9
Rate of STAT independent production of SLBO mRNA	m_β^o	0.03
Rate of desgradation of SLBO	δ_β	1
Rate of degradation of mRNA of SLBO	δ_{m_β}	0.3
Degradation rate of mRNA to SLBO due to miRNAs	$\delta_{B\sigma}$	0.5
Binding rate of STAT to slbo and dissociation rate	k_β^f	1
	k_β^b	6
Rate that slbo transitions into repressed state and rate it transitions out	$k_{\beta R}^f$	2
	$k_{\beta R}^b$	0.522

TABLE 1 Initial parameters that were used to establish bistability in the simplified model. These parameters were perturbed in order to gauge the sensitivity of the bistability of the system. Here, A and B represent APT and SLBO, respectively. Furthermore, the fraction of transcriptionally inactive apt and slbo genes have been written as α and β .

Fig 8 shows that multiple steady states emerge as STAT increases. Specifically for intermediate STAT activation we have three inter-sections. Through stability analysis (data not shown), we find that

we have two stable steady states, separated by one unstable steady state. Here, the left side of the steady state equation is plotted using a dashed gray line and the right side is plotted in intermediate level of STAT (black, solid), a lower level of STAT (gray, solid), and with an increased level of STAT (gray, dotted). At a low level of STAT activation, there is only one intersection corresponding to low SLBO production and lack of cell motility. At high STAT levels the system can only be at the activated level of SLBO production and the cell will move. Importantly, for intermediate STAT levels, three intersections emerge, where the system could be either activated or inactivated depending on the initial state of the system. This result is consistent with a model in which there is a threshold, as we predicted.

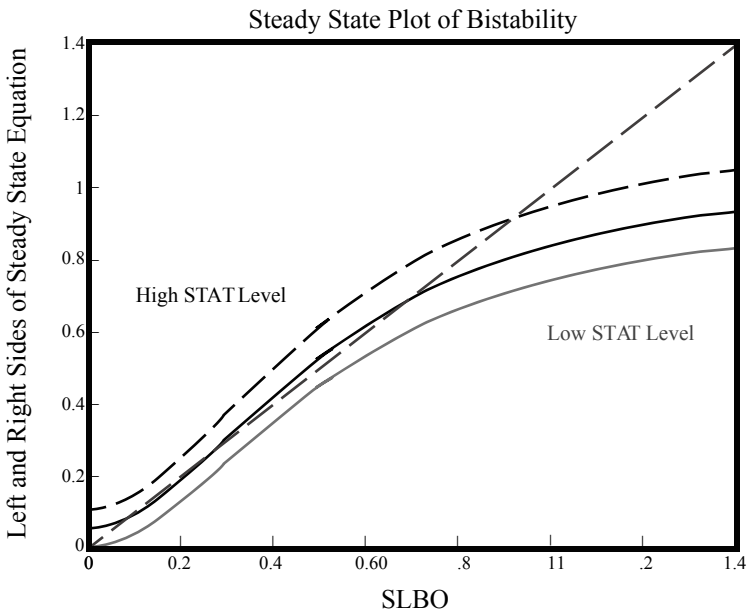


FIG. 8 A steady-state plot reveals bistability at an intermediate level of STAT activity. The left and right hand sides of the steady state equation plotted against SLBO for three levels of STAT activity. A high level of activated STAT causes only activation of the pathway, while a low level of activated STAT leads to only inactivation. However, if there is an intermediate level of STAT, the system could be either activated or inactivated.

This critical result shows that our reduced mathematical system of cross repression is sufficient to produce bistability. Since this model is based on the mechanisms that control the pathway, it helps to validate the observed biological result of bistability in the system. It reveals that this framework of simple cross repression is capable

of producing a bistable system, which is significant for our system and possibly for many others across the biological sciences as well, since many other systems are controlled by biologically similar mechanisms.

We continued our mathematical investigation by analyzing how sensitive the bistability of our model is to variation of the parameters. We wanted to see how the system responds to small changes to the parameters and how bistability was maintained. We did this, in part, because many of these parameters were not well established in previously published literature. We used, as a guideline, parameters from Harris et al. and Starz-Gaiano et al. to estimate several of our parameters. In Fig 8 STAT (which initially is equal to 1) is varied by 15 percent. This leads to a loss of bistability. By changing each parameter and noting how the endpoints of the bistable range move, we can get a sense of how flexible the system is.

We performed a local sensitivity analysis to quantify this change using two measures, 'Bistable Start' and 'Bistable Range' following the methodology of Harris et al (2011). 'Bistable Start' is the lowest STAT value that produces multiple equilibria and is a measure of how the left endpoint of the bistable domain moves in response to the perturbation. 'Bistable Range' is the measure of the change in the bistable range (distance from the left endpoint to the right endpoint) in response to the parameter perturbation. We computed these values for each parameter in the simplified system. Fig 9 shows these values. For example, k_A , the production rate of APT translation, shows a sizeable shift in the initial point of bistability to higher STAT and a significant increase in the range of bistability, while the degradation of APT, δ_A , shows a sizeable shift in the initial point of bistability to lower STAT and a significant decrease in the range of bistability. The reverse is true for the translation and degradation of SLBO. In general it appears that parameters related to enhancing APT increase the range of bistability and begin bistability at higher levels of STAT, while parameters related to enhancing SLBO tend to decrease the range of bistability and begin bistability at lower STAT levels. Furthermore, it is less critical to have exact parameter, given the range of values that can exhibit bistability.

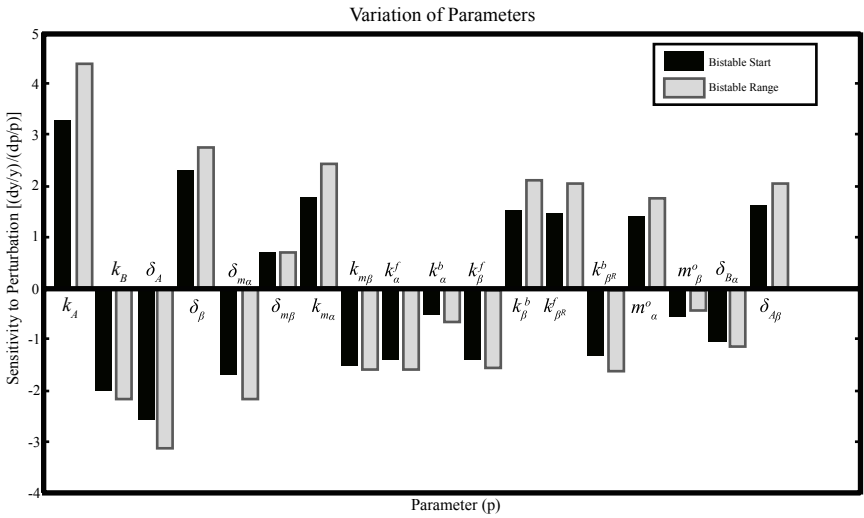


FIG. 9 A quantification of the sensitivity of bistability to parameter perturbation. Bistable Start is $[(R_1 - R_0)/R_0]/(\Delta p/p)$, where R_0 is the starting right limit point, R_1 is the right limit point after perturbation and p and Δp are the parameter value and change in parameter value (10%) respectively. Bistable Range is $[(R_1 - L_1) - (R_0 - L_0)]/(R_0 - L_0)/(\Delta p/p)$, where L_0 and L_1 are the starting and perturbed left limit point respectively. They inform where the range of bistability begins and how it responds to perturbation.

Ultimately, we have shown that the downstream effectors of the JAK/STAT pathway are capable of producing a bistable switch in cell fate. This result, coupled with the results from the bistability assay, confirms that the current model contains all of the important components. We also showed that with the simplified framework must have cooperativity of SLBO to repress APT in order to obtain bistability in the mathematical model. This observation leads us to believe that the full model likely needs to satisfy the same condition in order to achieve bistability, hence the B^2 in equation (9) of the full system.

Several testable hypotheses resulted from the mathematical analysis. One hypothesis is that cooperativity is likely in the repression system. One way cooperativity could occur is by the dimerization of two SLBO monomers or by the auto-regulation of *slbo* transcription, which could be tested in biological assays. Another finding is that the rate of translation is an important component in bistability. The parameter sensitivity analysis showed that the production of APT protein is particularly sensitive, which leads us to consider which biological components could affect this production. We hypothesize that one or more microRNAs, which can control translational rate,

are important. Using biological techniques we are working to determine if there are microRNAs that mediate the negative interactions between APT and SLBO.

MICRO-RNAS

WHILE THE INITIAL biological circuit shown in Figure 2A demonstrates the basic relationships among the factors, in reality, the molecular interactions are much more complicated. In fact, we suspect that the inhibitory pathways are regulated by intermediates as well. We believe that there is a possibility that SLBO may be inhibiting APT by upregulating the expression of a microRNA (miRNA) that targets the *apt* message. miRNAs are short RNA sequences of about 22 base pairs that are normally encoded within the genome. They bind to the 3' untranslated regions (UTRs), which are the regulatory regions, of mRNAs, and block protein translation by either inhibiting the translational process itself or by degrading the mRNA. Whether a given mRNA is degraded or translationally inhibited depends on the number of complementary base pairs between the microRNA and the mRNA itself (Enright et al., 2003). We believe that this type of molecule may play an important role in mediating the cross repression of APT and SLBO because both APT and SLBO are transcription factors. In fact, APT is known to inhibit STAT via upregulation of a miRNA (Yoon et al., 2011).

We performed a database search to determine candidate microRNAs (Betel et al 2010.; Betel et al., 2008; Enright et al., 2003). We took advantage of the fact that the entire genome of *Drosophila melanogaster* is sequenced and annotated in detail, as well as the fact that the DNA binding site consensus sequence is known for both SLBO and APT. Thus, we were looking for any miRNAs that were complementary to either the *slbo* or *apt* message, and in particular those that also had upstream sequences that matched the consensus binding site for the reciprocal protein. Since the mechanism by which SLBO represses APT is unknown, we are particularly interested in candidates upregulated by SLBO that bind to the 3' UTR of the *apt* mRNA. Strikingly, we found 14 candidates that matched our criteria. This strongly supports the idea that cross-repression between SLBO and APT is very important and therefore highly regulated. The summary of this data is shown in Table 2. Based on available reagents, microRNAs 87 and 284 will be the focus for further studies.

microRNA	Binds to	Number of SLBO binding sites	Number of APT binding sites
miR-962	Slbo 3' UTR	1	-
miR-9c	Slbo 3' UTR	1	(1*(9bp))
miR-9b	Slbo 3' UTR	-	(1*(9bp))
miR-927	Slbo 3' UTR	2	(1*(8bp))
miR-315	Slbo 3' UTR	3	2
miR-274	Slbo 3' UTR	-	-
miR-964	Slbo 3' UTR	2	-
miR-1002	Slbo 3' UTR	4	-
miR-252	Slbo 3' UTR and apt 3' UTR	2	2
miR-970	Apt 3' UTR	1	-
miR-87	Apt 3' UTR	1	1
miR-8	Apt 3' UTR	4	-
miR-284	Apt 3' UTR	1	1
miR-1	Apt 3' UTR	-	-
Let-7	Apt 3' UTR	3	-

TABLE 2. Putative microRNA that mediate the cross repressional relationship between SLBO and APT. The top half of the table shows the microRNAs that bind to the 3' untranslated region of the *slbo* mRNA, and the bottom half of the table shows the microRNAs that bind to the 3' untranslated region of the *apt* mRNA. The microRNAs 87 and 284 are the intermediates currently under investigation.

The current mathematical model can accommodate that microRNAs may play a role in the negative relationship between APT and SLBO. In fact, the mathematical formalisms can be virtually the same. We showed that for the reduced system to have bistability there likely is cooperativity of SLBO to repress APT at the gene level. If microRNAs do play a role in this system then it is straightforward to incorporate into the mathematics, but the result from the minimum bistability model will still hold. Based on the cooperativity of SLBO in suppressing APT supported by our mathematical model, we might even predict that SLBO would dimerize to initiate their transcription.

Essential biological processes are highly regulated, and often with redundant mechanisms that ensure robustness. We have used a combination of mathematical and biological strategies to dissect a molecular switch that sets a threshold between two different cellular behaviors. Our work has revealed many levels of regulation that contribute to this system, and we have established a set of

minimal required components to establish bistability. This minimal cross-repression system requires one of the components to act with cooperativity, which we can examine further *in vivo*.

In addition to the minimal system, the full system can also display bistability, and it is likely that both systems contribute to the outcomes *in vivo*, accentuating output of the entire network. We also showed that the physical removal of an activating signal via migration is not required for bistability, however it may provide a backup mechanism to help regulate the overall system. Finally, since the minimal bistability model indicated the importance of cross-repression, we searched for molecules that could mediate this regulation. We identified many new miRNA candidates that likely have a role in the specification of motile cells, and these will be a focus of further investigation.

Mathematical modeling of bistable systems has been of interest in many aspects of biology. The JAK/STAT signaling pathway, likewise, is important in many contexts, and is well known for a role in stem cell maintenance. Many of the components we have studied in the ovary are also essential in stem cells in other fly tissues, and we hope to extend our modeling analysis to provide insight to these other situations.

ACKNOWLEDGMENTS

The authors thank the members of the Starz-Gaiano laboratory for experimental assistance and members of the Undergraduate Biology/Mathematics program (UBM@UMBC) for helpful discussions and support. We used the following databases in our work: <http://www.microrna.org/microna/home.do> and <http://flybase.org/>. We also acknowledge the following funding sources: NSF-UBM to DS, SG, BEP, XG (PIs Dr. Leips and Dr. N. Neerchal); NSF-Career to MSG; UMBC Start-Up funding to BEP and MSG.

REFERENCES

- Arbouzova, N.I., and Zeidler, M.P. (2006). JAK/STAT signalling in *Drosophila*: insights into conserved regulatory and cellular functions. *Development* 133, 2605-2616.
- Baeg, G.H., Zhou, R., and Perrimon, N. (2005). Genome-wide RNAi analysis of JAK/STAT signaling components in *Drosophila*. *Genes Dev* 19, 1861-1870.
- Beccari, S., Teixeira, L., and Rorth, P. (2002). The JAK/STAT pathway is required for border cell migration during *Drosophila* oogenesis. *Mech Dev* 111, 115-123.
- Betel, D., Koppal, A., Agius, P., Sander, C., and Leslie, C. (2010). Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol* 11, R90.
- Betel, D., Wilson, M., Gabow, A., Marks, D.S., and Sander, C. (2008). The microRNA.org resource: targets and expression. *Nucleic Acids Res* 36, D149-153.
- Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). MicroRNA targets in *Drosophila*. *Genome Biol* 5, R1.
- Ghiglione, C., Devergne, O., Georgenthum, E., Carballes, F., Medioni, C., Cerezo, D., and Noselli, S. (2002). The *Drosophila* cytokine receptor Domeless controls border cell migration and epithelial polarization during oogenesis. *Development* 129, 5437-5447.
- Harris, R.E., Pargett, M., Sutcliffe, C., Umulis, D., and Ashe, H.L. (2011). Brat promotes stem cell differentiation via control of a bistable switch that restricts BMP signaling. *Dev Cell* 20, 72-83.
- Jang, A.C., Starz-Gaiano, M., and Montell, D.J. (2007). Modeling Migration and Metastasis in *Drosophila*. *J Mammary Gland Biol Neoplasia*.
- Keener, J., and Sneyd, J. (1998). *Mathematical Physiology*, Vol 8 (New York, NY, Springer).
- Montell, D.J. (2008). Morphogenetic cell movements: diversity from modular mechanical properties. *Science* 322, 1502-1505.
- Montell, D.J., Rorth, P., and Spradling, A.C. (1992). slow border cells, a locus required for a developmentally regulated cell migration during oogenesis, encodes *Drosophila C/EBP*. *Cell* 71, 51-62.
- Naora, H., and Montell, D.J. (2005). Ovarian cancer metastasis: integrating insights from disparate model organisms. *Nat Rev Cancer* 5, 355-366.
- Reiter, L.T., Potocki, L., Chien, S., Gribskov, M., and Bier, E. (2001). A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. *Genome Res* 11, 1114-1125.
- Silver, D.L., Geisbrecht, E.R., and Montell, D.J. (2005). Requirement for JAK/STAT signaling throughout border cell migration in *Drosophila*. *Development* 132, 3483-3492.
- Silver, D.L., and Montell, D.J. (2001). Paracrine signaling through the JAK/STAT pathway activates invasive behavior of ovarian epithelial cells in *Drosophila*. *Cell* 107, 831-841.
- Starz-Gaiano, M., Melani, M., Meinhardt, H., and Montell, D. (2009). Interpretation of the UPD/JAK/STAT morphogen gradient in *Drosophila* follicle cells. *Cell Cycle* 8, 2917-2925.
- Starz-Gaiano, M., Melani, M., Wang, X., Meinhardt, H., and Montell, D.J. (2008). Feedback inhibition of Jak/STAT signaling by aptonic is required to limit an invasive cell population. *Dev Cell* 14, 726-738.
- Xi, R., McGregor, J.R., and Harrison, D.A. (2003). A gradient of JAK pathway activity patterns the anterior-posterior axis of the follicular epithelium. *Dev Cell* 4, 167-177.
- Yoon, W.H., Meinhardt, H., and Montell, D. (2011). miRNA-mediated feedback inhibition of JAK/STAT morphogen signalling establishes a cell fate threshold. *Nat Cell Biol* in press/online.

“OH LUCY!”

Sitcoms and the 1950s Housewife in the Early Cold War

kelly jordan

Kelly Jordan is a Presidential Scholar pursuing a B.A. in History. She is currently a member of the Humanities Living Learning Community and plans to study abroad at Charles University in Prague, Czech Republic for the Spring 2012 semester. She is a junior anticipating graduation in May 2013, after which she plans to attend graduate school to pursue a master's degree in public history. She would like to thank Dr. Meredith Oyen of the Department of History and Laura Marshallsay, recent UMBC Graduate School History alumna, for their guidance on this paper from the initial brainstorming to the final draft.

Although I have always loved older situation comedies such as *I Love Lucy* and *Father Knows Best*, they began to interest me academically during Dr. Meredith Oyen's course HIST 201: Introduction to the Study of History, which required a research paper. Dr. Oyen and I hypothesized that the portrayal of women and families in 1950s sitcoms provided a model that Americans followed, perhaps due to the instability caused by the United States' involvement in the Cold War. To investigate this theory, I sought the opinions of both historians and sociologists through journal articles and scholarly monographs. I supplemented this research by reviewing episodes of various sitcoms and period advertisements. Interested in pursuing the topic further, I worked with Dr. Oyen over the summer to incorporate into this paper additional sources and perspectives from the historiography of the 1950s.

INTRODUCTION

WITH HIS TRADEMARK exclamation, “Ai-yai-yai-yai-yai,” Ricky Ricardo expresses his frustration with his well-meaning wife Lucy in the popular 1950s American sitcom *I Love Lucy*.¹ Although the Ricardos have provided countless hours of laughter for many viewers, a deeper truth lies in the midst of the exploits and ensuing comedy. Lucy, whose short-sightedness often places her in difficult situations, is an aspiring actress who also juggles her duties as a housewife and mother. Lucy constantly tries to add more excitement to her life, auditioning for commercials and attempting to bring her entrepreneurial dreams to fruition. When her plans fail, Lucy remains upbeat and hopeful, an example to her fellow housewives who may be experiencing the same difficulties as she. Despite the constant quibbling between Lucy and Ricky, their television marriage and parenthood provide a safe haven for them and an escape from the troubles of the outside world, especially from the Cold War tensions and fear of nuclear war. This is a consistent theme in situation comedies of the fifties; middle-class housewives often faced trying situations and dilemmas, but they ultimately found solace within strong, steady homes. The portrayal of women in sitcoms during the 1950s offered a model of idealized female roles for families in the insecurity of early Cold War America. This is evidenced by the inherent homogeneity in television, the pro-American sentiment in reaction to worldwide tensions, and the ensuing societal expectations of housewives.

AMERICAN HOMOGENEITY IN THE FACE OF COLD WAR INSECURITY

THE INSTABILITY OF the world arena played a role in the everyday life of the typical American family during the 1950s. Andrew J. Falk, an associate professor of history at Christopher Newport University, notes the following:

There were close ties between nation and family in the 1950s. Both required corporate organization, sturdy leadership, and obedient citizenship. Both also required vigilance to ensure security. As television producer Norman Lear recalled, the televised family of the 1950s was portrayed “as a protective shield, warding off external forces. Home was happy, secure, serene. If tempests arose,

they would ultimately be calmed there as well, and nothing ever suggested that security could fail.”²

Against the backdrop of the world’s turmoil, the home became a place of safety and control. Uncertainty and American prosperity combined in a unique way, as young people sought solace in the creation of their own families. As middle-class Americans became more prosperous, young people married earlier in order to pursue their own successful lives.³ Like Ricky and Lucy Ricardo of *I Love Lucy*, many working-class Americans worked hard to achieve a middle-class lifestyle, eventually moving from their city apartments to fast-growing suburbia.⁴ Television producers favored the safety and comfort of homogeneity; the media depicted almost solely white, Protestant, middle-class families with practical, tenderhearted stay-at-home mothers and hardworking, respectable fathers. The 1950s are today remembered as a time of happiness and conformity.⁵

As a reaction to the sense of discomfort from the Cold War, many Americans did their best to conform to societal expectations as a demonstration of their patriotism and loyalty or perhaps because of fear of public dishonor.⁶ Historian Stephanie Coontz observes, “A ‘normal’ family and vigilant mother became the ‘front line’ of defense against treason; anti-communists linked deviant family or sexual behavior to sedition.”⁷ To prevent any accusations or feelings of anti-American activities within their households, mothers kept a close watch on the behaviors of their children. In fact, American society often valued good citizenship over other personal assets and achievements, considering it to be nearly synonymous with morality.⁸

Additionally, the American population viewed communism as the antithesis of the new American way; former President Herbert Hoover’s 1958 bestseller, *Masters of Deceit*, decried extreme leftist thinking as “a way of life; a false, materialistic, ‘religion.’ It would strip man of his belief in God, his heritage of freedom, his trust in love, justice and mercy.”⁹ Indeed, world instability and the rising fear of communism encouraged American families to seek refuge within themselves.¹⁰ In contrast to communism, sitcom producers showed that housewives such as June Cleaver of *Leave it to Beaver*, Harriet Nelson of *The Adventures of Ozzie and Harriet*, and Margaret Anderson of *Father Knows Best* each owned a gleaming, immaculate kitchen, complete with a well-stocked pantry and equipped with the latest that technology had to offer. This encouraged audiences to “become model consumers and, by extension, model families.”¹¹

Mary Beth Haralovich explains this phenomenon by portraying the 1950s housewife as the center of the capitalist system, instrumental in the development of marketing and new merchandise.¹² Sitcoms of the era portrayed the commercial ideal of the family, but perhaps they also inadvertently stigmatized the very women to whom they catered, limiting the domain of the average housewife to her work around the home, despite the empowerment of women during World War II just a few years before.¹³

“AMERICANISM” AS A REACTION TO FEAR

AS THE TYPICAL family changed with the times, often due to fear of communism or prosecution for anti-American sentiments, the institution of pro-American values led to governmental changes as well. The Federal Communications Commission (FCC) established new guidelines in an effort to promote patriotism and the morals of capitalist society. Their Blue Book strongly encouraged producers to demonstrate their support of widespread American ideals and to broadcast shows that were in the “public interest” or else lose support for program renewal.¹⁴ Victor Pickard notes that the Blue Book “took the unprecedented – and unrepeated – step of making the privilege of holding broadcast licenses contingent upon meeting substantive public interest requirements.”¹⁵ American policymakers saw television as a powerful means of influencing the public’s paradigm, but worried that without guidance, this new medium could be used by those with anti-American sentiments to sabotage the appearance of American nationalism in the international arena.¹⁶

The FCC and other government organizations and sponsors quickly established a system to eliminate communist and other anti-American propaganda. This system limited the available television shows to those with vivaciously pro-American themes, thus presenting an idealized model of American life to a captivated audience. Sociologist Ella Taylor acknowledges, “Benign, consensual domestic comedies, such as *Leave It to Beaver*, *The Adventures of Ozzie and Harriet*, and *Father Knows Best*, emerged in this production climate from the motto of the successful network careerist – ‘least objectionable programming’....these shows proposed family life as a zany, conflict-free adventure.”¹⁷ In a time when foreign tensions ran deep, the FCC and other government organizations sought to prevent such tensions from arising at home. These new shows provided Americans with

a model of how a family should function, thus preserving stability in a difficult time.

Television shows of the 1950s exemplified socially-expected values, particularly with regard to the ideal family structure. Historian Stephen J. Whitfield observes that “Americanism,” or unwavering devotion and loyalty to the United States, stood in direct opposition to the perceived corruption of communism, and television reflected how “[the] belief system that most middle-class Americans considered their birthright—the traditional commitment to competitive individualism in social life, to the liberal stress on rights in political life, and to private enterprise in economic life—was adapted to the crisis of the Cold War.”¹⁸ Sitcoms of the era modeled symbols of pro-American sentiments, such as patriotism and civic pride.¹⁹ “Lesson in Citizenship,” an early *Father Knows Best* episode that first aired in October of 1954, portrayed the mindset of many Americans during the 1950s. Television father Jim Anderson, played by the actor Robert Young, finds his three children being unproductive and ignoring their mother’s call for them. With their mother looking on approvingly, he counsels them to be good citizens by responding and volunteering to help more eagerly, explaining that “active participation is the basis of good citizenship.”²⁰ Despite the lighthearted inflection of the show, the message remains clear: in order to demonstrate their patriotism, citizens should be actively involved in their communities. As Falk notes, the mid-1950s marked a “shift toward ‘living room lectures’....Wise parents on *Ozzie and Harriet*, *Father Knows Best*, and *Leave It to Beaver* defused the anger of burgeoning juvenile delinquents with ‘a good hardy talk.’”²¹ During the mid-1950s and especially the late-1950s, the sitcom family demonstrated cohesive serenity and growth as everyone settled into idealized familial roles.²²

Similar situations demonstrate consistent themes in 1950s sitcoms, focusing on patriotism and expectations of good citizens, in particular the female roles associated with patriotism and citizenship. In an *I Love Lucy* episode titled “Drafted,” Lucy Ricardo and Ethel Mertz open their husbands’ mail from the local military base and mistakenly believe that their husbands have been called up into the army.²³ Rather than allow themselves to be upset about their husbands’ upcoming departure, Lucy and Ethel busy themselves making socks for their brave men and consoling themselves by reminding each other that Ricky and Fred are fulfilling their patriotic responsibilities. This exemplifies their

supportive roles of the dutiful wives. Even when it is revealed that the men are not truly being drafted but are instead performing for soldiers at the military base, Ethel and Lucy acknowledge the bravery of the military men and the need for them to serve in any way that they are able.²⁴

Displays of patriotism were not only limited to support for civic duties and the military, but also evidenced in the lifestyles of middle-class American families. Historian Elaine Tyler May explains, “Materialism became dually important to the American family as a status symbol: within individual families, conformity to the new commercial ideals showed an anti-communist dedication to the American way of life; collectively, American prosperity became important in the ideological war against communism, particularly in the U.S.S.R.”²⁵ The television set found a place among many material status-symbols, becoming more affordable than ever in the 1950s. Vance Packard explains the importance of the television in *The Status Seekers*, published in the 1950s: “the TV aerial has symbolic significance in some areas of the nation....Many two-room shacks have thirty-foot towers above them.”²⁶ Magnavox advertised its televisions in a 1950 edition of *Time* as “a lasting investment in pleasure as well as a magnificent expression of...good taste and sound sense of values.”²⁷ One of its competitors, DuMont, also linked television ownership with family values in its advertisements. Its 1950 ad in *The Saturday Evening Post* proclaims, “There is great happiness in television... great happiness in the home where the family is held together by this new common bond – television.”²⁸ The widespread ownership of television sets and the relatively small selection of television shows allowed featured programs to become influential, creating the American perspective of the ideal 1950s family.²⁹ Shows such as *Leave It to Beaver* mixed idealism with reality; Haralovich notes that “[s]triving for verisimilitude, the stories were based on the ‘real life’ experiences of the scriptwriters in raising their own children.”³⁰ Such families displayed specific gender roles that did not demonstrate much individuality: as an expression of their patriotism, women acted, dressed, and felt a certain way. Each member of the typical American family had a part to play, and although deviation from societal norms was possible, radical thinkers who dared to attempt such a rebellion would risk social or familial ostracism.³¹

SOCIETAL EXPECTATIONS AS A REACTION TO AMERICANISM

COMMERCIALISM FURTHERED THESE familial roles, as the increased income of many middle-class families led to an increase in consumer optimism and department store sales. These department stores catered to the changing desires of middle-class women. In 1958, *Time* noted, “More secure in their personal income, consumers are now planning stepped-up purchases, particularly in housing and household goods.”³² Cold War-era situation comedies, particularly during the 1950s, portrayed a happy home as an indicator of success. The characters of June Cleaver (*Leave It to Beaver*) and Margaret Anderson (*Father Knows Best*) were middle-class suburban mothers whose lives were supposedly made easier by their modern kitchen and home appliances. Mary Beth Haralovich notes:

*The suburban home and consumer products have presumably liberated Margaret and June from the domestic drudgery which marks Alice’s daily existence. A major portion of the comedy of The Honeymooners’ (1955–1956) working-class urban family is derived from Ralph and Alice Kramden’s continual struggle with outmoded appliances, their lower-class taste, and the economic blocks to achieving an easy assimilation into the middle-class through homeownership and the acquisition of consumer goods.*³³

Although working-class families found themselves less able to afford newfangled appliances, their plight may not have been much different than that of the middle-class families’. New time-saving devices supposedly gave homemakers freedom; however, women spent more time on their housework during the 1950s, even with new appliances and the availability of convenience food.³⁴ Additionally, the amount of time spent on childcare doubled from the 1920s to the 1950s.³⁵

Purchases of brand-name conveniences were a way for American housewives not only to demonstrate their success, but also to exhibit their loyalty and Americanism. The United States displayed its technological advances proudly to the communist Union of Soviet Socialist Republics (USSR) in an attempt to intimidate the USSR. American Vice President Richard M. Nixon boasted that the United States had developed a suburban model home that working-class

families could afford, where technology and science provided innovative appliances intended to allow homemakers a more leisurely lifestyle.³⁶ Housewives of period sitcoms, such as Alice Kramden of *The Honeymooners*, applauded the purchase of commodities, encouraging families to “live above [their] means – the American way.”³⁷ May expounds upon this sentiment: “Commodities would solve the problem of the discontented housewife, foster pride in the provider whose job offered few intrinsic rewards, and allow children to ‘fit in’ with their peers.”³⁸

Increased incomes led to increased purchasing, and with that came more intense advertising. American society during the 1950s encouraged women to exude femininity and competence, displaying their prosperity through their appearance. Both advertisers and sitcom producers used the media to demonstrate societal expectations. Women sought youthfulness and femininity, striving to maintain their mystique with the hope of attracting successful men. Product promoters “put the libido back into advertising,” encouraging women to use their sex appeal to secure their futures.³⁹ *Father Knows Best* demonstrates such expectations: “In one episode...Margaret is dressed in dungarees, sweatshirt, and loafers, her hair covered by a scarf as she scrubs paint from her youngest daughter, Kathy. When Betty witnesses this sight, she laughs, ‘If you aren’t a glamorous picture!’ As Jim arrives home early, Betty counsels Margaret, ‘You can’t let [your husband] see you like this!’”⁴⁰ Society expected housewives to maintain a calm, put-together appearance, even with their families, in order to show prosperity and little deviance from the expected norms. The appearance of conformity was important to the female societal role during the 1950s, but this conformity often came at the price of concealing deep-seated issues within housewives and their families.⁴¹

Strict societal expectations of functionality encouraged family responsibility. The establishment of new behavioral and ideological norms left little room for arguments, rebellion, or even abuse. Ella Taylor posits that Americans emphasized family responsibility for problems within society, blaming familial conflicts and abuses on a lack of cohesion in family life.⁴² Because the media, including sitcoms, valued and publicized homogeneity and normality as the epitome of the American ideal, families felt encouraged to hide their problems.⁴³ Eugene Rodney, who produced *Father Knows Best*, was quoted in *Cosmopolitan*, where he “identified the program’s audience as the middle-class and middle-income family. ‘It’s people in that

bracket who watch us. They don't have juvenile delinquent problems. They are interested in family relations, allowances, boy and girl problems."⁴⁴ In a 1956 edition of *Cosmopolitan*, Kenneth Rhodes praised *Father Knows Best* for its "polite, carefully middle-class, family-type entertainment, possibly the most non-controversial show on the air waves."⁴⁵ Even while the actors and actresses of such sitcoms struggled with alcoholism and marital stress in their personal lives, their characters remained steadfast in their dedication to societal norms and their families.⁴⁶

Because family members often kept their personal problems to themselves, the home front seemed quiet and serene, and housewives appeared to enjoy their work at home. For the most part, women seemed to complete their duties cheerfully and willingly, serving their families by supporting their husbands, raising their children, and keeping up the appearance of clean, happy, Americanized households. As Coontz explains:

Nineteenth-century middle-class women had cheerfully left housework to servants, yet 1950s women of all classes created makework in their homes and felt guilty when they did not do everything for themselves...By the mid-1950s, advertisers' surveys reported on a growing tendency among women to find "housework as a medium of expression for...[their] femininity and individuality."⁴⁷

Studies suggested that the majority of American homemakers were content with their decision to devote their lives to their families. Dr. E. Lowell Kelly of the University of Michigan sought to explore such contentedness by conducting the Kelly Longitudinal Study, in which he surveyed three hundred New England couples starting in the late 1930s, with the majority of his in-depth research taking place in 1955.⁴⁸ By asking questions such as "What has marriage brought you that you could not have gained without your marriage?" and "What did you have to sacrifice or give up because of your marriage?" Kelly further deciphered the phenomenon of the new American family, which "believed that affluence, consumer goods, satisfying sex, and children would strengthen their families, enabling them to steer clear of potential disruptions."⁴⁹ Many of the interviewed women expressed a sense of fulfillment from their family responsibilities and viewed their happiness as exclusive to their marital and family experiences.⁵⁰ May points out that such women did not convey

unhappiness with their decision to marry rather than pursue more autonomous life choices: “One said marriage gave her a ‘happy, full, complete life; children; a feeling of serving some purpose in life other than making money.’ Another remarked, ‘I’m not the “career girl” type. I like being home and having a family....Working with my husband for our home and family brings a satisfaction that working alone could not.’”⁵¹ In the end, society expected housewives to be cheerful, hardworking, put-together, and vigilant for their families. Although seemingly content with such complacency, by the end of the decade many housewives had tired of their lifestyle, feeling that it held them back from a truly fulfilling life.

GROWING DISCONTENT WITH CONFORMITY

WHILE MANY WOMEN seemed content at home, Betty Friedan questioned the reality of this sentiment in her renowned work, *The Feminine Mystique*: “Consider, as a symptom, the increasing emphasis on glamour in the women’s magazines: the housewife wearing eye makeup as she vacuums the floor – ‘The Honor of Being a Woman.’ Why does ‘Occupation: housewife’ require such insistent glamorization year after year?”⁵² The 1950s middle-class lifestyle provided housewives with a specific role to play in the family performance, for truly such conformity must be an act. However, Friedan stated that, by the end of the 1950s, women found themselves growing tired of the lack of individuality.⁵³ May notes that many advertisers promoted consumerism as a “means for assimilation into the American way of life,” preventing, or perhaps simply postponing, a true societal realization of the reality of 1950s culture.⁵⁴ Although they may not have encountered pressures associated with deviation from the norm, housewives privately expressed their frustrations with the seemingly unending housework and childcare.⁵⁵ By the end of the decade, however, many women had become more comfortable with expressing their discontentment. Friedan quips, “American women’s unhappiness is merely the most recently won of women’s rights.”⁵⁶ While women struggled to maintain their happiness, regardless of how shallow it truly was, the sitcom mother remained as calm and collected as ever. Margaret Anderson, the quintessential 1950s housewife, did not complain or express any hint of discontentment.⁵⁷ This new housewife-mother, a hard worker dedicated to the happiness of her family, embodied feminine grace.⁵⁸

Although the put-together appearance of the sitcom housewife is even today one of the most iconic memories of the 1950s, it may have been an elaborate *façade*. Situation comedies modeled the ideal life for American families in the hope that it would protect capitalism and other pro-American standards. The media depicted a disregard for the individuality of women, as evidenced by the homogeneity inherent in many sitcoms, advertisements, and surveys, as well as the rebellion against such norms that exploded at the end of the decade. The character of Lucy Ricardo remained upbeat regardless of whatever mischief or shortcoming she faced, a positive and enthusiastic example for housewives at home. Such optimism and even suppression of individuality was not as genuine in American society as it was on television. Given the idealized model of the American housewife provided in sitcoms, which was likely a threat to the happiness of housewives in the 1950s, perhaps “Ai-yai-yai-yai-yai” is right.

ENDNOTES

- 1 Marc Daniels, "Men Are Messy," *I Love Lucy* (CBS, December 3, 1951).
- 2 Andrew Justin Falk, *Upstaging the Cold War: American Dissent and Cultural Diplomacy, 1940-1960* (Amherst: University of Massachusetts Press, 2010), 171.
- 3 Stephanie Coontz, *The Way We Never Were: American Families and the Nostalgia Trap* (New York: Basic Books, 2000), 26, 263.
- 4 *Ibid.*, 76.
- 5 Falk, *Upstaging the Cold War*, 172–173. Although the vast majority of situation comedies of the 1950s centered around white Protestant families, there were a few shows based on the lives of ethnic and racial minorities, such as *Beulah* (1950 to 1953), *Life with Luigi* (1952 to 1953), and, the most popular of these, *The Goldbergs* (1949 to 1954). These shows, however, featured minorities in what was assumed to be their "natural habitat," and even *The Goldbergs* became less popular when producers tried to move the Jewish working-class family from the Bronx to the suburbs. Falk notes on page 173 that such sitcoms "also suggest the Cold War tendency to forgo progressive notions of race and ethnicity in favor of the familiarity and nostalgia that stereotypes offer."
- 6 Mary Beth Haralovich, "Sitcoms and Suburbs: Positioning the 1950s Homemaker," *Quarterly Review of Film & Video* 11 (May 1989): 61–62.
- 7 Coontz, *The Way We Never Were*, 33.
- 8 Stephen J. Whitfield, *The Culture of the Cold War* (Baltimore: Johns Hopkins University Press, 1996), 65.
- 9 *Ibid.*, 68.
- 10 *Ibid.*
- 11 Ella Taylor, "From the Nelsons to the Huxtables: Genre and Family Imagery in American Network Television," *Qualitative Sociology* 12 (Spring 1989): 4.
- 12 Haralovich, "Sitcoms and Suburbs," 61.
- 13 *Ibid.*, 62.
- 14 Falk, *Upstaging the Cold War*, 132.
- 15 Victor Pickard, "The Battle Over the FCC Blue Book: Determining the Role of Broadcast Media in a Democratic Society, 1945-8," *Media Culture & Society* 33 (March 23, 2011): 172.
- 16 Falk, *Upstaging the Cold War*, 7.
- 17 Taylor, "From the Nelsons to the Huxtables," 16.
- 18 Whitfield, *The Culture of the Cold War*, 53.
- 19 *Ibid.*
- 20 William D. Russel, "Lesson in Citizenship," *Father Knows Best* (CBS, October 10, 1954).
- 21 Falk, *Upstaging the Cold War*, 171–172.
- 22 *Ibid.*, 171.
- 23 Marc Daniels, "Drafted," *I Love Lucy* (CBS, December 24, 1951).
- 24 *Ibid.*
- 25 Elaine Tyler May, *Homeward Bound: American Families in the Cold War Era* (New York: Basic Books, 2008), 17.
- 26 Vance Packard, *American Social Classes in the 1950s: Selections from Vance Packard's The Status Seekers* (Boston: Bedford/St. Martin's, 1995), 60–61.
- 27 Print advertisement for Magnavox combination television-radio-phonograph unit, "Prized Possession in Smart Young Homes," (1950) found on the Duke University Library Digital Collection Online, http://library.duke.edu/digitalcollections/adaccess_TV0151/#info, [accessed May 9, 2011].
- 28 Print advertisement for the DuMont Tarrytown television set, "Home Team," (1950) found on the Duke University Library Digital Collection Online, <http://library.duke.edu>.

- edu/digitalcollections/adaccess_TV0456, [accessed May 9, 2011].
- 29 Falk, *Upstaging the Cold War*, 71.
- 30 Haralovich, "Sitcoms and Suburbs," 63.
- 31 *Ibid.*, 18.
- 32 "State of Business: Consumer Optimism," *Time*, December 8, 1958, <http://www.time.com/time/magazine/article/0,9171,864545,00.html>, [accessed May 9, 2011].
- 33 Haralovich, "Sitcoms and Suburbs," 64.
- 34 Coontz, *The Way We Never Were*, 27.
- 35 *Ibid.*
- 36 May, *Homeward Bound*, 162.
- 37 *Ibid.*, 172.
- 38 *Ibid.*
- 39 Betty Friedan, *The Feminine Mystique* (New York: W. W. Norton & Company, 1963), 225.
- 40 Haralovich, "Sitcoms and Suburbs," 18.
- 41 Friedan, *The Feminine Mystique*, 225.
- 42 Taylor, "From the Nelsons to the Huxtables," 8.
- 43 Coontz, *The Way We Never Were*, 31.
- 44 Haralovich, "Sitcoms and Suburbs," 64.
- 45 *Ibid.*, 62.
- 46 Falk, *Upstaging the Cold War*, 172.
- 47 Coontz, *The Way We Never Were*, 27.
- 48 May, *Homeward Bound*, 11–12.
- 49 *Ibid.*, 11–12, 28–29.
- 50 *Ibid.*, 29.
- 51 *Ibid.*
- 52 Friedan, *The Feminine Mystique*, 65.
- 53 *Ibid.*, 24.
- 54 May, *Homeward Bound*, 172.
- 55 *Ibid.*, 173.
- 56 Friedan, *The Feminine Mystique*, 24.
- 57 Russel, "Lesson in Citizenship."
- 58 Friedan, *The Feminine Mystique*, 59.

ANTIBIOTIC RESISTANCE IN ESCHERICHIA COLI

Characterization of ketolide-resistant mutants

asmara qamar

Asmara Qamar is a 2010 UMBC graduate with a B.S. in Biological Sciences and a B.A. in English Literature. She will be attending the University of Maryland School of Dentistry in the fall of 2012. She would like to thank her advisors Dr. Janice Zengel and Dr. Lasse Lindahl, her mentor Dr. Sephorah Zaman, the members of the Zengel/Lindahl lab, and her mom for their guidance, help, and support. She would also like to thank the Office of Undergraduate Education for funding her project through an Undergraduate Research Award.

This paper discusses the characterization of novel *Escherichia coli* mutant strains that are resistant to the antibiotic cethromycin. I started this project in the Zengel/Lindahl lab in 2009. I was drawn to this lab because of my interest in molecular biology. This lab primarily studies ribosomes, so I took up the project there to isolate *E. coli* strains that were resistant to the novel antibiotic cethromycin, a ribosome-targeting drug. Research comes with successes and failures, and initially, although I identified a few interesting mutants, I was not able to predict their patterns of growth and could not freeze them away. Mistakes teach, and soon after recognizing their growth patterns, I managed to freeze away four mutants and characterize them. Although there is much about the molecular aspects of ribosome-mediated antibiotic resistance that is unknown and yet to be uncovered, this paper gives insight into the character of mutations in the ribosomal proteins L4 and L22 that confer resistance to the antibiotic cethromycin.

INTRODUCTION

THE RIBOSOME IS the living cell's factory for protein synthesis. A bacterial ribosome contains two subunits: the small 30S subunit, and the large 50S subunit. After DNA is transcribed to mRNA, the ribosome, using tRNAs to interpret the mRNA code, travels along the mRNA and strings together specific amino acids to form a peptide chain that eventually folds into a fully formed protein (Steitz, 2008).

Macrolide antibiotics, which include erythromycin, are widely prescribed. Erythromycin-derivatives, such as cethromycin, constitute a novel class of antibiotics called ketolides (Schlunzen *et al.*, 2003). Macrolides and ketolides target the ribosome by binding to a specific target in the RNA component of the 50S subunit at the entrance of the peptide exit tunnel, a passage from which the newly translated peptide chain emerges. This binding hinders the progression of the nascent peptide chain and thereby blocks protein synthesis (Schlunzen *et al.*, 2003; Steitz, 2008; Schlunzen *et al.*, 2001).

The chemical structure of cethromycin, a ketolide, is shown in Figure 1a, and erythromycin's structure is shown in Figure 1b. Ketolides derive their name from the presence of a ketone on C3 which holds the cladinose sugar in erythromycin. Groups on cethromycin which are not present on erythromycin, such as a cyclic carbamate attached to the lactone ring, and a quinoline at an allylic position attached to the lactone ring through a flexible linker, may allow cethromycin to make more interactions and bind with a greater affinity to the ribosome than erythromycin. Cethromycin makes more contact points with 23S rRNA compared to macrolides. In addition to the contact with 23S rRNA domain V which erythromycin exhibits, cethromycin also makes contacts with domains II and IV (Schlunzen *et al.*, 2003). Indeed, cethromycin has been shown to bind with twenty fold higher affinity than erythromycin to the *Haemophilus influenzae* ribosome (Cao *et al.*, 2004).

Cethromycin interacts with 23S ribosomal RNA through hydrogen bonds with nucleotides (in the *E. coli* numbering system) A2058, A2057, C2611, U2609, and U790, and hydrophobic interactions with nucleotides A2059, C2610, G2505, A2062 and U1782. Additionally, this binding is strengthened by van der Waals forces and spatial fitting between the antibiotic and the surrounding rRNA (Schlunzen *et al.*, 2001).

Although the ribosome is made mostly of RNA, it also contains proteins that are involved in ribosomal structure and assembly,

among other functions. Ribosomal proteins L4 and L22 extend from the surface of the ribosome to its core; elongated loops of these proteins form the narrowest constriction of the lining of the peptide exit tunnel. Ribosome-binding antibiotics such as macrolides do not physically interact with ribosomal proteins. However, mutations in the loops of these proteins have been shown to confer resistance to the macrolide erythromycin in *E. coli* (Zaman *et. al*, 2007). This result is most likely due to the effect of these proteins' presence on rRNA proximal to the antibiotic binding site (Schlunzen *et al.*, 2001). This study has determined that similar mutations in these L4 and L22 also confer resistance to cethromycin. However, the exact character of these mutations and extent of resistance granted by them differs subtly between macrolides and ketolides.

MATERIALS AND METHODS

Determination of minimum inhibitory concentration

WILD TYPE *E. COLI* strain AB301 was grown in Luria broth (LB) overnight at 37°C with shaking. Aliquots (100 µl) of the overnight culture were plated on LB agar plates containing 25, 50, 100, 150 or 200 µg/ml cethromycin and monitored for growth after 3-5 days of incubation at 37°C. To see if there was any difference in growth inhibition in liquid media, 10 µl of an AB301 culture was added to 2 ml tubes of LB containing 5, 10, 15 and 30 µg/ml cethromycin, and incubated at 37°C overnight.

Mutant selection

AB301 IN LB was grown at 37°C with shaking overnight. Aliquots (100 µl) of this culture were then plated on 30 µg/ml cethromycin LB agar plates, and the plates were incubated for 3-5 days at 37°C. Single resistant colonies were purified by re-streaking on fresh LB agar plates containing 30 µg/ml cethromycin.

To make freezer stocks, two methods were used. For relatively fast growing mutants, single colonies were used to inoculate one ml of LB containing 30 µg/ml of cethromycin and grown overnight at 37°C; 500 µl of this culture was added to 500 µl of 50% glycerol and frozen at -80°C. Since all strains grew slower in liquid media than on agar places, exceptionally slow growing mutants such as L22-N93E and L22-77Δ12 were first purified, then several colonies were picked

from an LB + cethromycin plate and thoroughly vortexed in 500 μ l LB + 30 μ g/ml cethromycin and 500 μ l 50% glycerol, and frozen at -80°C .

Sequence analysis

SINGLE COLONIES OF resistant strains were used for PCR amplification, using primers 01818 (for the L4 gene) and 01820 (for the L22 gene). The conditions for amplification were as follows: denaturing was set at 95°C , with ten minutes for the first cycle and one minute for the remaining; annealing was set at 50°C ; and elongation was set at 72°C for one minute, and ten minutes for the last cycle. Agarose gel electrophoresis was used to monitor the success of the PCR amplification. The Qiagen PCR purification kit was used to purify the DNA. Sequencing was done by PCR using big dye reaction mix. After the PCR amplifications, samples were kept out of light at all times. The samples were cleaned by ethanol precipitation by incubating in 125 mM EDTA and ethanol. They were then spun at 13K, and the pellet was washed with 70% ethanol. The supernatant was removed by tipping the tubes horizontally, disrupting and aspirating the supernatant with a pipette, absorbing the remaining liquid with a kim wipe, and then drying the pellet in a speed vacuum for 20 – 25 minutes.

Growth of mutants on cethromycin and erythromycin

ERYTHROMYCIN-RESISTANT L22 and L4 mutants isolated by Zaman *et al.* (2007) were plated on LB agar plates containing 50, 75 and 150 μ g/ml of cethromycin, and on agar containing the minimum inhibitory concentration of erythromycin (150 μ g/ml), as determined by Zaman (2007). The four strains isolated for cethromycin resistance, L22-93/6, L4-66/2, L22-N93E and L22-77 Δ 12, were tested for cross resistance by plating on 150 μ g/ml of erythromycin. Their degree of resistance was tested by plating them on LB agar containing 30, 50, 75, 100, 150 and 200 μ g/ml cethromycin.

Measuring the growth rate of mutant strains

A PATCH OF mutant colonies from LB agar and 30 μ g/ml cethromycin plates was used to inoculate LB containing 30 μ g/ml cethromycin. AB301 was inoculated to LB. These cultures were shaken at 37°C ; samples were withdrawn at various times during growth of the strains and optical densities were measured at 450 nm. The growth was plotted versus time on a natural log scale.

β -galactosidase induction assay

STRAINS L22-93/6, L4-66/2, L22-N93E and L22-77 Δ 12 were grown overnight in liquid LB + cethromycin, and AB301 was grown overnight in LB. These cultures were used to inoculate 15 ml of AB minimal media, supplemented with glycerol, thiamine, and casamino acids, to a starting OD 450 nm of about 0.05, and grown at 37°C with shaking. Once the cultures reached an O.D. of 0.25 - 0.4, they were induced by addition of 0.5 M IPTG, giving a final concentration of 1 mM, and 200 μ l samples were withdrawn at various times after induction. The samples were added to 800 μ l Z buffer, 50 μ l 1% deoxycholate, 10 μ l chloramphenicol and 40 μ l toluene, all kept at 0°C, and samples were then kept in 37°C for 45 minutes to evaporate the toluene and ethanol (solvent for chloramphenicol). ONPG was added to the samples after transferring them to a 37°C water bath. After 12 - 15 hours, the reaction was stopped by the addition of 1 M sodium bicarbonate, and the absorbance of the samples was measured against an un-induced sample at 550 and 600 nm. β -galactosidase units were calculated using the equation described in Miller (1972), and plotted against time of sample withdrawal. The lag time was extrapolated by fitting a line through the values obtained by the most reliable spectrophotometer readings.

Ribosome preparation

400 ML OF LB was inoculated to a starting OD 450 of 0.05 with an overnight culture containing LB and cethromycin, and incubated with shaking at 37°C. When the culture reached an OD of 1.5-2, cells were harvested by pelleting them at 8000 rpm, and washing the pellet with buffer A (20 mM HEPES-KOH pH 7.5, 6 mM MgCl₂, 30 mM NH₄Cl, 6 mM β -mercaptoethanol). Cells were lysed in the French press at 1000 psi. To remove insoluble material, the lysate was incubated for 15 minutes at 0°C with 15 μ l RNase-free DNase. The supernatant was transferred to a MLA80 ultracentrifuge tube, and then spun at 50,000 rpm for 4 hours in an ultracentrifuge. The supernatant was decanted, and the pelleted ribosomes were resuspended in 200 μ l buffer A by gentle rocking overnight in a 4°C coldroom.

Erythromycin binding assay

2.5 A²⁶⁰ UNITS OF ribosomes, 20 μ l of 1 μ Ci/ml C¹⁴ erythromycin, 20 μ l 5X erythromycin binding buffer (250 mM Tris HCL, pH 7.8,

80 mM Mg-acetate, and 300 mM KCL) and water up to 100 μ l were incubated for 15 minutes in a 30°C water bath. The samples were chilled on ice, and then passed through a nitrocellulose membrane on a Millipore aspirator by washing three times with approximately 3 mls of erythromycin wash buffer (25 mM Tris HCL, pH 7.8, 20 mM Mg-acetate, and 60 mM KCL in 30% ethanol). The nitrocellulose filters were dried and immersed in scintillation liquid and passed through a Beckman-Coulter LS 6500 scintillation counter at 10 minutes of counting time per sample. The cpm from mutant ribosomes were adjusted to exclude background noise (the noise level was obtained by running a sample without ribosomes through the counter), and normalized to the wild type ribosomes' signal.

RESULTS

Minimum inhibitory concentration

ON LB CETHROMYCIN plates, growth of AB301 was inhibited on concentrations higher than 25 μ g/ml cethromycin. In liquid LB, no significant growth could be observed above 5 μ g/ml cethromycin. On LB plates, 30 μ g/ml cethromycin inhibited wild type but allowed the proliferation of a few resistant colonies. Thus, it was used as the concentration for selection.

Mutant selection

ELEVEN MUTANTS WERE identified by sequencing, and four were further characterized. These mutations are summarized in Table 1. The sequences of L22 and L4 mutants are shown in Figure 2a and b. Their locations on L4 and L22 are shown in Figure 3a and b. Seven mutant strains were lost due to growth insufficient for isolation, unexpectedly poor growth when transferred directly to liquid media, and technical errors. Once it was observed that mutant strains grew more poorly in liquid media than in agar, freezer stocks were made by the alternative method described above in Materials and Methods (Mutant selection). Figure 4 shows the strains L22-93/6, L4-66/2, L22-N93E and L22-77 Δ 12 growing in LB plates containing 30 μ g/ml cethromycin.

All eleven mutations either mapped directly within the tentacle loops of L4 and L22, or affected the area through frameshift muta-

tions. There were roughly equal numbers of L4 and L22 mutants, although the L22 mutations were more drastic.

Five L4 mutations were identified; all mapped between codons 62 and 68, a small region at the tip of the tentacle, which also, as Figure 5 shows, is spatially closest to the antibiotic binding site. The missense point mutation Q62L substitutes leucine for glutamine at codon 62; R67C substitutes cysteine for arginine at codon 67; and A68E substitutes glutamate for alanine at codon 68. The mutation 65 Δ 9 deletes the amino acids threonine, glycine and arginine from the tip. The insertion 66/2, affecting the same part of the tentacle, is a direct repeat of the preceding codons 65 and 66, thereby adding a threonine and a glycine to the loop.

L22 mutations were more scattered than L4 mutations, although all affected the loop. 77 Δ 12 is a deletion of twelve bases starting from codon 77; this removes the amino acids aspartate, glutamate, glycine, proline and serine. Although this deletion removes nucleotides within its initial and last codons, 77 and 81 respectively, the first base of codon 77 and the last two bases of codon 81 combine to introduce a glycine to the code.

A93E is a point mutation substituting glutamate for alanine at codon 93, and at the same codon the mutation 93/6 inserts four amino acids. This insertion results from a direct repeat of the four preceding codons, 90-93, inserting the amino acids lysine, glycine, arginine and aspartate.

71 Δ 7 is a frameshift deletion of seven bases ranging from codons 71 to 73 which introduces a premature stop signal at codon 77. The mutation removes a valine, threonine and an adenine, and changes the downstream sequence to read for a lysine, phenylalanine, serine and stop, from the original lysine, isoleucine, phenylalanine and valine. Figure 6a shows a prediction of the truncated protein that might result from this mutation. 78 Δ G is the deletion of a guanine at codon 78 and introduces a premature stop codon at codon 82; a predicted truncated protein that might result from this mutation is shown in Figure 6b. This mutation changes the originally following glutamate, glycine, proline, serine, and methionine to lysine, alanine, arginine, alanine and stop.

The mutation 49 Δ 11 is another out-of-frame deletion of eleven base pairs, ranging from codons 49 to 53, which introduces a premature stop signal at codon 57. A predicted truncated protein is shown in Figure 6c. This mutation removes a lysine (codon 49), valine, leu-

cine, glutamate and serine, and rearranges the reading frame after codon 49 to code for the amino acids asparagine, cysteine, histidine, cysteine and stop, from the original alanine (codon 54), isoleucine, alanine and asparagine.

Although mutants 71 Δ 7, 78 Δ G, and 49 Δ 11 were identified, their slowness and lack of growth did not allow their isolation or characterization.

Growth of mutants on cethromycin and erythromycin

FIGURES 7A-D SHOW known erythromycin-resistant L4 and L22 mutants plated on 150 μ g/ml erythromycin, and on 50, 75 and 150 μ g/ml cethromycin. The canonical mutants N281 and N282 displayed sparse growth on cethromycin at the concentration of 50 μ g/ml (data not shown). On the whole, erythromycin-resistant strains with point mutations such as L4-G66S, G66D, G66R, Q62K and N282 showed little to no resistance to cethromycin, although G66C was resistant to up to 150 μ g/ml cethromycin. Strains with insertions or deletions showed relatively more resistance than point mutants. L22-93/6, L4-66/2, L22-N93E and L22-77 Δ 12, all grew on 150 μ g/ml erythromycin, as shown in Figure 7e. Table 2 lists the maximum concentrations of cethromycin on which growth of the characterized mutants was observed. L22-93/6 was the most tolerant and exhibited growth on up to 200 μ g/ml cethromycin. L4-66/2, L22-77 Δ 12 and L22-N93E grew in up to 50, 50 and 30 μ g/ml cethromycin respectively.

Growth rates of selected mutant strains

GROWTH RATES OF isolated strains are shown in Figure 8a. The growth rate of AB301 in LB is shown in Figure 8b. The doubling times of the four mutants in LB + 30 μ g/ml cethromycin are listed in Table 2. L22-93/6 had the shortest doubling time at 47 minutes, followed by L4-66/2 at 90, L22-77 Δ 12 at 110, and L22-N93E at 120 minutes.

IPTG induced β -galactosidase elongation lag time

ISOPROPYLTHIO-D-GALACTOSIDASE (IPTG) induces the *lacZ* gene, causing it to transcribe for the enzyme β -galactosidase. The time between this induction and the appearance of the β -galactosidase enzyme provides information about how long ribosomes are taking to translate the β -galactosidase mRNA. β -galactosidase induction assay plots are shown in Figure 9. The lag time for AB301

can be noted as approximately 95 seconds, 130 seconds for L22-77 Δ 12, 170 seconds for L22-93/6, 110 seconds for L4-66/2, and 130 seconds L22-N93E.

Erythromycin binding assay

THE NORMALIZED BINDING affinity of erythromycin to ribosomes isolated from the mutant strains is listed in Table 2. The ribosomes of L4-66/2 bound at 4% the affinity of the wild type; L22-93/6 at 45%, L22-N93E at 90%, and L22-77 Δ 12 at 66%.

DISCUSSION

THE MINIMUM INHIBITORY concentration for cethromycin, at 30 μ g/ml, proved lower than that determined for erythromycin's 150 μ g/ml. This result corresponds with the notion that cethromycin's additional contact points and flexibility of attached groups allow it to bind with more affinity to the ribosome. Previously isolated erythromycin-resistant mutants also showed lesser growth in cethromycin than in erythromycin. Among erythromycin-resistant mutants, point mutants were generally less resistant to cethromycin, and among the newly isolated cethromycin-resistant mutants, the point mutant L22-N93E had the least resistance and the slowest doubling time.

Insertion and deletions among mutants granted relatively higher tolerance for cethromycin. The two insertions, L22-93/6 and L4 66/2 were both direct repeats of upstream sequences, possibly duplicated due to RNAP II slippage during transcription. The reduced resistance of point mutants to cethromycin compared to insertion or deletion mutants may suggest that point mutations disturb interactions specific to a certain antibiotic, while insertions and deletions change the constriction of the exit tunnel in a dramatic enough way to allow interruption of a broad range of antibiotic-ribosome interactions. For example, the erythromycin-resistant point mutants isolated by Zaman *et al.* (2007) are also resistant to a number of other macrolide antibiotics, such as spiramycin and tylosin – all similar antibiotics. The lack of resistance of these mutants to cethromycin suggests a specific range of the point mutations.

Among the spontaneously occurring mutations isolated by Zaman *et al.* (2007), more L4 mutations were found to confer resistance to erythromycin than L22 mutations. This trend did not follow

with cethromycin, where the frequency of resistance-enabling L22 mutations roughly equaled if not exceeded the frequency of L4 mutations. This difference in proportion may suggest that cethromycin interacts with rRNA bases that are closer to L22. L22 mutations have also been implicated in non-structural roles towards resistance, and it is possible that mutations affecting these non-structural roles allow for resistance against cethromycin (Moore *et al.*, 2008). Although if a structural argument is followed, the locations of the resistance-granting mutations on L22 correlate with its position along the peptide exit tunnel. Two of the mutations in L22, which were both isolated, are in codon 93, near the tip of the elongated loop. Figure 5 gives an idea of how close the tip of L22's loop is to bases A2058 and A2057, the former of which is the primary binding site for macrolides and ketolides (Novotny *et al.*, 2004). Herein also lies the only spontaneously occurring point mutation identified in L22, suggesting that even minor changes in this critically positioned part of the protein can cause significant changes in the geometry of the surrounding rRNA, and consequently in the interactions it is capable of.

Unlike L4, L22's elongated loop lies somewhat parallel with the exit tunnel, forming more of the tunnel's wall (Berisio *et al.*, 2003). Three of the identified mutations in L22 occurring between codons 70-80 are possibly responsible for changing or even removing the loop, and so may affect the lining of the exit tunnel. The deletion 77 Δ 12 removes 12 bases, while the relatively drastic frameshift mutations 77 Δ 7 and 78 Δ G introduce stop codons before the loop. The resultant proteins could either be incorporated as they are, or, due to their malformation, be degraded by proteases through proteolysis (Lodish, 2004). In that case, these mutations either remove the loop or the whole protein. The fact that they occur in a particular area suggests that they mainly remove the loop; however, their slow rates of growth, which hindered isolation, suggest perhaps that these malformed proteins were not properly contributing to the function and assembly of the ribosome, and therefore did not allow for appropriate ribosomal function and dependent cell growth. The deletion 49 Δ 11 is a frameshift mutation even further upstream than the aforementioned mutations, and results in a highly truncated peptide. This reduced protein may be digested by proteases.

L4 mutations enabling resistance to cethromycin were similar to L4 mutations enabling erythromycin resistance in that they were all concentrated at the tip of the loop. The lack of drastic spontaneous

mutations in L4 suggests that it may play important non-structural roles, most likely in assembly, that do not allow properly functioning ribosomes to form if L4 is severely malformed.

The doubling time for all mutants, as expected, was slower than that of the wild type. The mutant with the shortest doubling time as well as the highest resistance, L22-93/6, may have altered the tip of L22's loop in such a way as to significantly change the antibiotic-binding site. Another mutation in the same location, N93E, leads to the longest doubling time and the least resistance. This may be because cethromycin's flexibility in binding causes point mutations to fare poorly. However this cannot be said for certain since L4 point mutants were not characterized.

Also, as expected, all mutants had a longer IPTG induced elongation lag time. L22-93/6 had the longest lag time, which suggests that although the mutated ribosome tolerated high concentrations of antibiotic and still functioned well enough to allow a relatively low doubling time, it was changed in a way that interfered with the ribosomal response.

Although cethromycin's binding affinity to mutant ribosomes was not tested, it can be assumed, from their similarity of structure, that erythromycin and cethromycin have qualitatively similar binding affinities to the ribosome. Since the four cethromycin-resistant mutants showed resistance to erythromycin, it can be safely said that the mechanism through which resistance to cethromycin is enabled is also applicable to erythromycin-resistance. Overall the binding data of the four mutants followed the trend discussed by *Zaman et al.* (2007) where L22 mutants bind erythromycin, and L4 mutants do not. The L4-66/2 ribosome does not bind erythromycin, while the L22 mutated ribosomes in 93/6, N93E and 77 Δ 12 all bind erythromycin to certain extents. 93/6, however, does not bind as strongly as the other two L22 mutants, possibly because, as mentioned above, the loop-tip location of the mutation may cause changes in the surrounding rRNA. The point mutant N93E continues to bind the ribosome, raising a question as to how a potentially structure-changing mutation, occurring at the tip of the L22 loop, allows resistance while still allowing the antibiotic to bind.

Through the identification and characterization of cethromycin resistant mutants, this study confirmed that various mutations in the proteins L4 and L22 confer resistance to antibiotics that function by targeting the peptide exit tunnel. However, increased binding affinity of the antibiotic to the ribosome, and flexibility of binding that

allows attachment to the tunnel despite changes in the environment of the tunnel, leads to a comparatively more potent antibiotic such as cethromycin.

Gene	Location (codon)	Mutation	Strain name
L4	65 - 67	9 base pair deletion	L4-65Δ9
	62	CAG (gln) to CTG (leu)	L4-Q62L
	67	CGT (arg) to TGT (cys)	L4-R67C
	68	GCG (ala) to GAG (ghu)	L4-A68E
	66	6 base pair insertion	L4-66/2
L22	71-73	7 base pair deletion	L22-71Δ7
	49-53	11 base pair deletion	L22-49Δ11
	78	deletion of a guanine	L22-78ΔG
	77-81	12 base pair deletion	L22-77Δ12
	93	GCA (ala) to GAA (ghu)	L22-A93E
	93	12 base pair insertion	L22-93/6

TABLE 1 List of Novel Cethromycin Resistant L4 and L22 Mutants

Strain		Doubling time (min)	Maximum tested cethromycin resistance (μg/ml)	Maximum tested erythromycin resistance (μg/ml)	β-galactosidase induction lag time (sec)	Normalized Erythromycin binding affinity
Wt	AB301	30	5	>100	95	1
a7	L22-93/6	47	150	150	150	0.45
a20	L4-66/2	90	50	150	110	0.04
a3	L22-N93E	120	30	150	120	0.90
a13	L22-77Δ12	110	50	150	120	0.66

TABLE 2 Characteristics of Cethromycin Resistant Mutants and Wild Type Parent Strain AB301

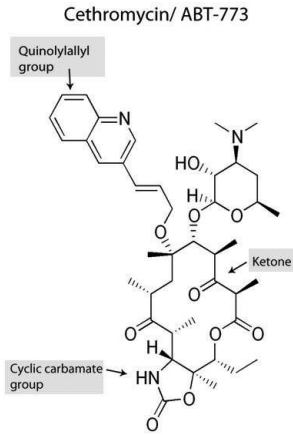


FIG. 1A. Chemical structure of cethromycin. Taken and edited from <http://en.wikipedia.org/wiki/File:Cethromycin.png>

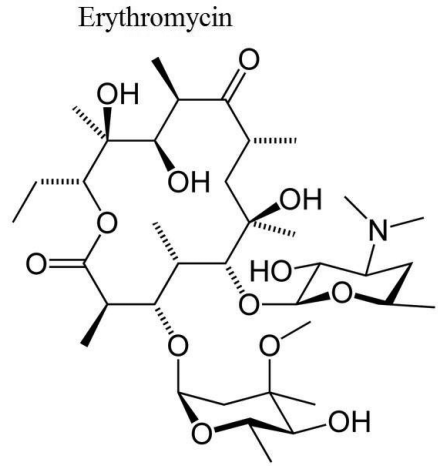


FIG. 1B. Chemical structure of erythromycin. Taken and edited from <http://en.wikipedia.org/wiki/File:Erythromycin-2D-skeletal.png>

L4 mutations

	61	62	63	64	65	66	67	68	69
wt	CGC	CAG	AAA	GGC	ACC	GGC	CGT	GCG	CGT
Q62L	CTG								
R67C							TGT		
A69E							GAG		
65 Δ9	CGC	CAG	AAA	GGC	ACC	GGC	CGT	GCG	CGT
	64	65	66			67	68	69	
66/2	GGC	ACC	GGC	ACC	GGC	CGT	GCG	CGT	

FIG. 2A Sequences of L4 mutations

L22 mutations

	70	71	72	73	74	75	76	77	78	79	80	81	82	83
wt	AAA	GTT	ACG	AAA	ATT	TTC	GTA	GAC	GAA	GGC	CCG	AGC	ATG	AAG
77 Δ12	AAA	GTT	ACG	AAA	ATT	TTC	GTA	GAC	GAA	GGC	CCG	AGC	ATG	AAG
71 Δ7	AAA	GTT	ACG	AAA	ATT	TTC	GTA	G						
78 ΔG	AAA	GTT	ACG	AAA	ATT	TTC	GTA	GAC	GAA	GGC	CCG	AGC	ATG	A
	49	50	51	52	53	54	55	56	57					
wt	AAA	GTT	CTG	GAA	TCT	GCC	ATT	GCT	AAC					
49 Δ11	AAA	GTT	CTG	GAA	TCT	GCC	ATT	GCT	AA					
	90	91	92	93	94	95	96							
wt	AAA	GGT	CGT	GCA	GAT	CGC	ATC							
A93E	AAA	GGT	CGT	GAA	GAT	CGC	ATC							
	90	91	92	93			94	95	96					
93/6	AAA	GGT	CGT	GCA	AAA	GGT	CGT	GCA	GAT	CGC	ATC			

FIG. 2B Sequences of L22 mutations. Premature stop codon indicated by underlining of blue bases.

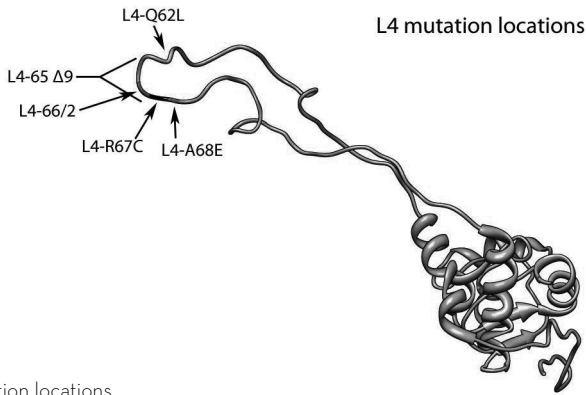


FIG. 3A L4 mutation locations

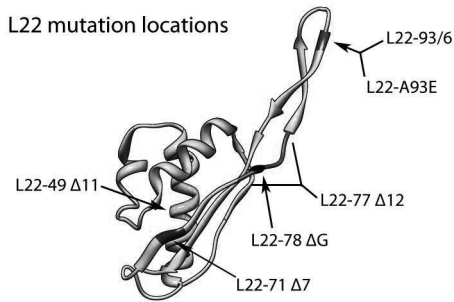


FIG. 3B L22 mutation locations

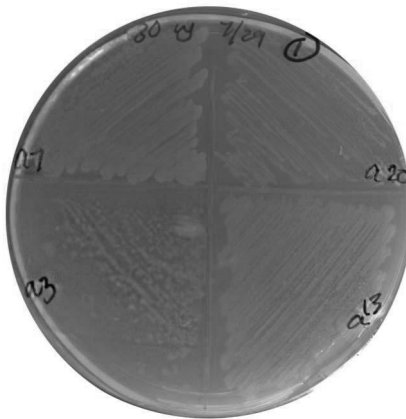


FIG. 4 Mutants L22-93/6, L4-66/2, L22-N93E and L22-77 Δ 12 on agar and 30 μ g/ml cethromycin. (Clockwise) A20 is L4-66/2, A13 is L22-77 Δ 12, A3 is L22-N93E and A7 is L22-93/6.

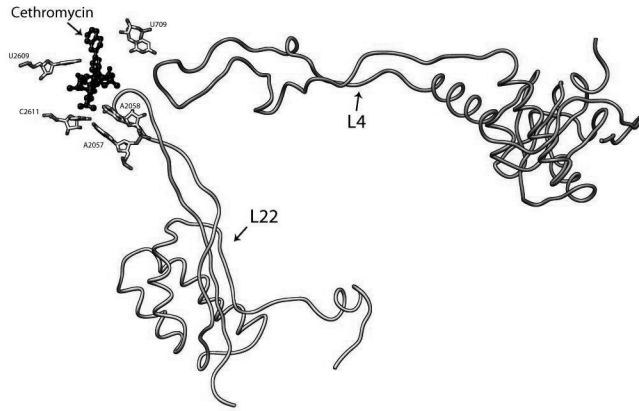


FIG. 5 Cethromycin's binding to the peptide exit tunnel and the relative positions of the proteins L4 and L22.

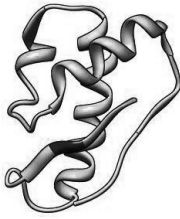


FIG. 6A Predicted truncated protein resulting from L22-71 Δ 7

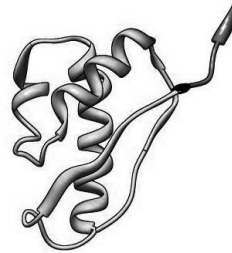


FIG. 6B Predicted truncated protein resulting from L22-78 Δ G

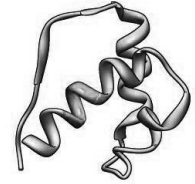


FIG. 6C Predicted truncated protein resulting from L22-49 Δ 11



FIG. 7A Erythromycin-resistant mutants on 150 μ g/ml erythromycin

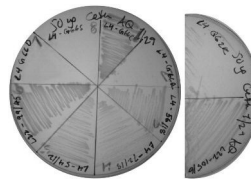


FIG. 7B Erythromycin-resistant mutants on 50 μ g/ml cethromycin

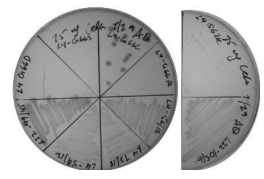


FIG. 7C Erythromycin-resistant mutants on 75 μ g/ml cethromycin



FIG. 7D Erythromycin-resistant mutants on 150 μ g/ml cethromycin



FIG. 7E L22-93/6, L4-66/2, L22-N93E and L22-77 Δ 12 on 150 μ g/ml erythromycin

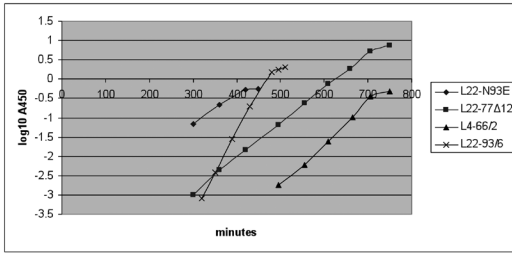


FIG. 8A Growth rates of L22-93/6, L4-66/2, L22-N93E and L22-77/12 on LB and 30 µg/ml cethromycin

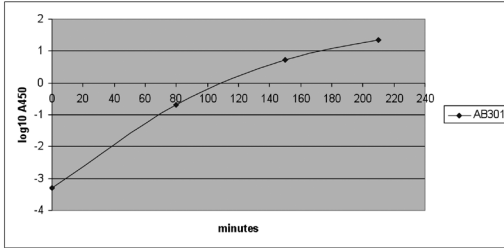


FIG. 8B Growth rate of AB301 on LB

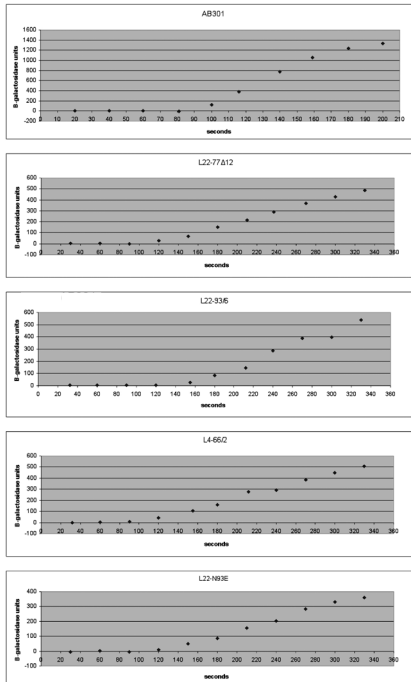


FIG. 9 IPTG induced B-galactosidase elongation lag time

REFERENCES

- 1 **Berisio, R., F. Schlunzen, J. Harms, A. Bashan, T. Auerbach, D. Baram, and A. Yonath.** 2003. Structural insight into the role of the ribosomal tunnel in cellular regulation. *Nat. Struct. Biol.* 10:366-370.
- 2 **Cao, Z.; Zhong, P.; Ruan, X.; Merta, P.; Capobianco, J.O.; Flamm, R.K.; Nilius, A.M.** (2004). Ribosome affinity and the prolonged molecular postantibiotic effect of cethromycin (ABT-773) in *Haemophilus influenzae*. *Int J Antimicrob Agents* 24:362-8.
- 3 **Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP, Zipursky SL, Darnell J** (2004). "3". *Molecular Cell Biology* (5th ed.). New York: W.H. Freeman and CO. pp. 66-72.
- 4 **Miller, J.H.** (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- 5 **Moore, S. D., and R. T. Sauer.** 2008. Revisiting the mechanism of macrolid-eantibiotic resistance mediated by ribosomal protein L22. *Proc. Natl. Acad. Sci. U. S. A.* 105:18261-18266.
- 6 **Novotny, G. W., L. Jakobsen, N. M. Andersen, J. Poehlsgaard, and S. Douthwaite.** 2004. Ketolide antimicrobial activity persists after disruption of interactions with domain II of 23S rRNA. *Antimicrob. Agents Chemother.* 48:3677-3683.
- 7 **Schlunzen, F., J. M. Harms, F. Franceschi, H. A. Hansen, H. Bartels, R. Zarivach, and A. Yonath.** 2003. Structural basis for the antibiotic activity of ketolides and azalides. *Structure* 11:329-338.
- 8 **Schlunzen, F., Zarivach, R., Harms, J., Bashan, A., Tocilj, A., Albrecht, R., Yonath, A. and Franceschi, F.** (2001) Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature*, 413, 814-821.
- 9 **Steitz TA** (2008) A structural understanding of the dynamic ribosome machine. *Nat Rev Mol Cell Biol* 9:242-253.
- 10 **Zaman, S., Fitzpatrick, M., Lindahl, L., Zengel, J.** (2007) Novel mutations in ribosomal proteins L4 and L22 that confer erythromycin resistance in *Escherichia coli* *Mol. Microbiol.* 66:1039-1050.

YOUNG MOTHERS IN A *QUILOMBO*
Praia Grande, Brazil

christina briscoe

Christina Briscoe graduated summa cum laude from UMBC in May 2011 with a B.S. from the Interdisciplinary Studies Program. Through this program, she created her own major, combining her passions for public health and anthropology. She received support through the UMBC Premier Award and the Honors College. During her study-abroad program in Brazil, she became interested in the reproductive health experience of young women in *quilombos*, the topic of the present study. In this study, she was guided by Dr. Damiana de Miranda from the School for International Training, Dr. Climene de Camargo from the Universidade Federal da Bahia, and Dr. Bambi Chapin at UMBC. She is especially grateful to the people of Praia Grande. In February 2012, with the support of a Fulbright Research Grant, she will return to that community to conduct further research on family planning. Following this, she intends to pursue a joint M.D./Ph.D. program in Medical Anthropology.

In January 2010, I enrolled in the study abroad program School for International Training (SIT) in Salvador da Bahia, Brazil in order to learn more about health care in Latin America. The program included a mentored research project in conjunction with the Universidade Federal da Bahia (UFBA). This paper examines the results from that exploratory, interdisciplinary research, principally discussing the characteristics and experience of young mothers in the *quilombo* Praia Grande, Brazil. During my fieldwork in that community, I conducted surveys with 18 of the 19 young mothers (13-19 years old) and in-depth, personal interviews with five of them. I also engaged in participant observation, the collection of a life history, and the facilitation of an interactive group encounter with the mothers. My field research, initially aimed at exploring the difficulties brought on by adolescent motherhood, led me to see the wide variety of individual mothering experiences and a conflict between young women's ideal and actual opportunities for schooling and employment. This study highlights the value of close, interpretative research to inform clinical and public health practices in marginalized communities.

INTRODUCTION

FROM 1550 TO 1850, Portuguese traders are estimated to have kidnapped and sold four to eleven million Africans into slavery in Brazil (Conrad, 1985; Klein, 1987). The large stretches of unclaimed land, due to the country's political and physical geography, promoted the formation of *quilombos*. These are defined as self-sustaining communities settled by escaped slaves or their descendants (Araujo, 2000). After the end of slavery in 1888, *quilombos* continued to form, with over 7,000 estimated to be in existence today (Caldas and Garcia, 2007; Freitas, 1981). Situated on the island of Ilha de Maré in the northeast of Brazil, Praia Grande is one of these *quilombos* and the site of my fieldwork from March to June 2010. The inequalities that I saw there inspired me to conduct research to increase the understanding and awareness of *quilombos* and the difficulties its marginalized citizens continue to face. Within the community, research with adolescent mothers also helped to highlight the structural violence experienced by Praia Grande residents.

Numerous preconceptions exist about adolescent pregnancy, such as it being an objective and universal public health problem. Many public health organizations claim that adolescent pregnancy contributes to the continuation of poverty and poor health outcomes for young women. However, these views on adolescent pregnancy are not valid in all parts of the world. In particular, narratives of the young mothers in Praia Grande illustrate how adolescent pregnancy and motherhood do not continue poverty in Praia Grande but rather are the response to poverty and other structural inequalities in this area. By highlighting the difference between these two interpretations of adolescent pregnancy, this paper ultimately focuses on decreasing misguided blame placed on young women for their difficult circumstances by those of different ethnic or class backgrounds. This research also stresses the importance of close ethnography and participant observation to inform effective and culturally respectful family planning care for young women. The results of a survey, a life history of one mother, and interviews with the young mothers of Praia Grande, along with supporting evidence from epidemiological and anthropological perspectives on adolescent pregnancy, are given below to support these conclusions.

LITERATURE REVIEW

PREVENTION OF ADOLESCENT pregnancy is a prominent public health agenda for organizations like the Center for Disease Control and Prevention (CDC), World Health Organization (WHO), and Sistema Unico de Saude (SUS, the Brazilian universal health care system), with the premise that adolescent pregnancy contributes to the continuation of poverty and poor health outcomes for young women. The following sections outline how this causal claim is not supported by the primary research and why it has become a generally accepted fact.

Epidemiological Perspectives: Inconsistencies in Literature on Adolescent Pregnancy

ADOLESCENT PREGNANCY IS correlated with low socio-economic status, low life-time wages, higher dropout rates, increased drug and alcohol abuse, higher inter-partner violence risk, higher rates of depression, and higher rates of child abuse (Levandowski et al., 2002). The close empirical relationships among these factors have helped to perpetuate the idea that if public health outreach could prevent women from having children young, the women would be able to complete their educations and perhaps be more likely to leave poverty. However, this conclusion erroneously equates correlation with causation.

A review of health outcomes specifically attributable to early motherhood, rather than due to *pre-existing* risks of women likely to become teen mothers, reveals that the conclusions of research on the topic are much more ambiguous and situational than much public health policy assumes. In her article “Teen Pregnancy is not a Public Health Problem,” Dr. Lawlor (Head of Division of Epidemiology, Department of Social Medicine, University of Bristol) noted, “Larger studies and those employing methods specifically designed to adequately control for confounding factors (for example using sibling comparisons) suggest that young age is *not* an important determinant of pregnancy outcome or of the future health of the mother” (Lawlor & Shaw, 2002, p. 552).

The importance of accounting for pre-existing differences between 10-19 and 20-29 year-olds who have children can be seen in d’Orsi & Carvalho’s (1998) research that analyzed 107,883 post-

partum outcomes in Rio de Janeiro. Initially, young maternal age was correlated with the poor outcomes; however, after controlling for residence zipcode (CEP) and socioeconomic class, the researchers found that adolescence was not a statistically significant predictor of infant/mother mortality, low birth weight, and Apgar score. Rather, *bairros periféricos*, or poor neighborhoods, as well as low levels of access to health care and education were primary risk factors (d'Orsi & Carvalho, 1998).

There are a few key points from epidemiological literature on adolescent pregnancy that must be stressed:

- 1) Maternal/infant biophysical health risks: No study that appropriately adjusted results for socioeconomic and health-care access differences between groups of women through multivariate statistical analysis has shown significantly elevated biological risk for the mother or fetus after the maternal age of fifteen (Barker & Castro, 2002; Berenson et al., 1997; Bukulmez & Deren 2000; Conde-Agudelo et al. 2005; Costa et al., 2002; Fraser et al. 1995; Gama et al., 2001; Jorgensen, 1993; Levandowski et al., 2008; Lubarsky et al. 1994; Olausson et al. 1997; Oliveira, 1998; Ribeiro et al., 2000; Satin et al. 1994; Smith & Pell, 2001; Yazlle et al., 2002).
- 2) Cycle of poverty: The purportedly negative socioeconomic and affinitive consequences of young motherhood are inconsistently found and depend primarily on context. While it seems reasonable to conclude that motherhood during adolescence has a negative impact on scholastic achievement independent of other factors (Card & Wise, 1978; Paraguassú et al., 2005), that impact does not seem to be as dramatically negative as generally thought. Wage gaps are slight, with women who became mothers during adolescence sometimes having the higher average salary (Hotz et al., 2005).
- 3) Mother/infant relationship: Socioeconomic status and the perceived quality of kinship support seem to be more predictive than maternal age alone in determining the observed quality of mother/infant relationships (Folle et al. 2004; Luster & Mittlestaedt 1993; McAnarney et al. 1984; Osofsky et al. 1993).

Primary literature shows the public health policy of early pregnancy prevention to be divorced from substantive, epidemiological analysis and evidence (Lawlor, 2002). As a result, medical and public health professionals must carefully sidestep representing adolescent

pregnancy as pathological (Lincoln et al., 1976). Pregnancy is not a disease in young women but rather a reflection of different, pre-existing social and economic realities.

Anthropological Perspectives: Adolescence is Culturally and Historically Shaped

LITERATURE ON ADOLESCENCE and adolescent mothering shows them to be historically and socially variable in ways that make it clear that teen-motherhood is only problematic in some contexts. The characteristics currently associated in the industrialized West with adolescence — such as irresponsibility, emotional volatility, identity achievement, and the development of autonomy — emerged relatively recently. Arising first in the European, wealthy elite during the 17th century, current norms of adolescence spread to the working class largely after the Industrial Revolution, when work outside the home and education, rather than birth, began increasingly to determine social status (Adaszko, 2005). Before that time, women we now consider adolescents generally were expected to begin assuming their adult roles and responsibilities of beginning families (Merces Bahia Bock, 2004). Although the women's rights movement victories for education in the 20th century resulted in many female adolescents having increased opportunities for studies and preparation, the choice to delay other dreams made little sense for poor women who had few meaningful opportunities for a high-quality education and professional career.

According to Shelee Colen's theory of stratified reproduction, cultures tend to value some reproductive futures and roles over others. The dominant class determines the norms of fertility and reproduction in a way that best fits their progeny, although "women perform physical and social reproductive labor structured by economic, political, and social forces and differentiated unequally across hierarchies of class, race, ethnicity, gender, and place in a global economy" (Colen in Ginsburg & Rapp, 2007, p. 78).

Such norms that benefit the progeny of the dominant class may not fit the needs of others, who are nonetheless subjected to their scrutiny. For poor, young women living in Philadelphia, Eden & Kefalas (2005) found that pregnancy and motherhood were pragmatic and logical constructions of meaning; affinitive gains were high and economic lost potential was temporary or negligible. Similarly, Lock & Kaufert (2006) argue that poor, adolescent women's choices

to become pregnant should not be demonized or romanticized, but should be seen as a pragmatic response to what “they perceive to be in their best interest” (p. 481). Yet, though they may defy norms, the young women’s awareness of their failures to achieve societal ideals also forces them to see their personal victories of motherhood in the midst of poverty and violence through the eyes of others: a shameful failure (Eden & Kefalas, 2005).

The experience, embodiment, and practice of motherhood for adolescents on the ground may differ from the generally held ideals on fertility timing in the larger community (Eden & Kefalas, 2005; LeVine, 1988; Strathern, 1992 in Ginsburg & Rapp, 1995). As Barlow & Chapin (2010) argue, for mothers of all ages and from all places, intimate interpersonal relationships, micro-ecosystem, and women’s own personalities modify and interpret the larger meanings that circulate and are shaped by macro-scale forces like the market economy. For this reason, a close and sensitive observation of adolescent pregnancy in its local context is necessary in order to appropriately evaluate the effects of motherhood on young women.

Anthropological research aids our understanding of the complex relationships among adolescence, motherhood, medicine, economic development, and reproduction. Each are shaped, embodied, and enacted in different cultural settings. Through this interdisciplinary analysis on Ilha de Maré, this research attempts to provide such a nuanced, grounded analysis of young women’s experiences of motherhood.

METHODOLOGY

AT THE BEGINNING of this research, no count of the number of young mothers (ages 10 to 19) living in Praia Grande existed. In the initial phases of this research, I set out to identify women who met the study criteria of (1) being a female between the ages of 10 and 19, the ages for adolescence according to the World Health Organization, (2) having at least one living child, and (3) agreeing to participate in the research. All research was performed in concordance with Resolution 196/96 of the Brazilian National Council on Health for free and informed consent and with the approval of Dr. Climene Camargo (UFBA).

Following the community health worker (CHW) as she visited

homes in Praia Grande helped me to initially identify nine women who met the study criteria; one was unable to give informed consent. The public health post allowed access to the family records that had been compiled by the CHW of Praia Grande over the last 15 years. Through birth-date checking, I counted the number of young women (10-19 years old) currently living on the island and identified the households in which a woman (10-19 years old) lived with a child who was at least ten years younger than her. Despite the extended family structure, the CHW said that it was unlikely that a young child would not live with his or her mother. This approach allowed the successful identification of ten additional young mothers.

The survey, included in the appendix, served to identify socio-demographic characteristics of the young mothers, household and environmental contexts, and reproductive health indicators. This data allowed an initial evaluation of the situation of most young mothers and aided in the development of rapport and trust for later personal interview encounters.

While orally completing the surveys with the women, I also scheduled five semi-structured, personal interviews to discuss life-themes with the young mothers. These interviews were limited in their depth and quality because of the short time available for research on the island: many of the women were very busy with caring for their children, *mariscando* (shellfishing), or making the artisanal baskets that the men of the island took to sell in the market in Salvador every Saturday. One woman's life history, taken over the course of four days, elaborated on themes seen in the rest of the interviews and explained them more fully.

In order to condense some of the relevant topics and to include the perspectives of the women with whom I had not developed as strong of a relationship, I held a group session for eight participants of the study. The gathering served both to thank the women for participating and to initiate a rich conversation around their collective experience. Based on a qualitative analysis of the themes that had arisen in the interviews with the individual women, I posed questions to the group about (1) their feelings upon initially becoming pregnant and the reactions of those close to them, (2) the birth control methods they used, and (3) the resources they and their children lacked on the island. The group discussion was a particularly productive activity since the women's own interests and concerns guided the conversation, providing rich, nuanced qualitative data on their experiences.

RESULTS

Prevalence

OF PRAIA GRANDE'S 1,737 residents, 19 of the 161 (11.8%) young women (age 10-19) had at least one child at the time of study. Seven of the 12 women aged 20 (58.3%) in Praia Grande had at least one child. Two women under the age of twenty were pregnant on the island during the time of the study.

Participant Demographic Characteristics

THE AVERAGE AGES for conception and giving birth were 16.4 and 17.3 years old, respectively (range: 13-19). Most of the pregnancies occurred in what the World Health Organization considers late adolescence, after the age of fifteen, when cohort studies have shown that the biological risk of giving birth is no higher than for women in their early twenties (Levandowski, 2008). The average age of participants in the study was 18.9 years old. Ninety-four percent (n=17) of the young women who participated in the study became pregnant after fourteen years of age.

According to official statistics on race from the *Instituto Brasileiro de Geografia e Estatística* (IBGE), 95% of the inhabitants of Praia Grande are black (negro). However, all of the women in the study self-identified as *morena*, an ambiguous term used in Brazil that can indicate a wide range of skin colors and physical characteristics and that has little correlation to any particular official category used by the IBGE.

All but three of the study participants had lived on the island their entire lives. Two had lived for brief periods in other parts of the state of Bahia to study; one had moved from Salvador to the island to live with her husband.

Maternal Health

OF THE WOMEN, one-third (n=6) reported a health problem during the pregnancy. In all but one case, these problems involved weight gain, hypertension, or anemia. In the last case, the woman almost died during the birth, according to her "*porque o menino tava gordo demais para virar certinho*" (because the baby was too fat to turn properly). Seventeen percent (n=4) of the women had Caesarian sections, which is less than the Brazilian average (32%). All of the women gave birth in the maternity ward of a public hospital in Salvador.

More than three-fourths of the women reported completing at least seven pre-natal visits (average=7.8 visits) as recommended by Brazil's Ministry of Health. While the average number of visits was slightly higher for the women who completed their pre-natal care in the local, two-year-old health clinic on the island (n=8, average=8.3) than those who had to make the day long journey to the hospital (n=10, average=6.7), the difference was not statistically significant according to a chi square analysis. The small number of participants may be inadequate to reflect different usage trends and experiences in the hospital/clinical settings.

Marital Status

FIFTEEN OF THE women (83%) described themselves as *juntada* or joined with their partners. During the group session, the women collectively defined *juntada* as “married, just not on paper.” There was also one woman who was officially married in the church. Of the two who described themselves as single, one was in a relationship but did not live with her boyfriend. Thus, all but one of the participants described themselves as being in a stable, heterosexual, and monogamous relationship. The average number of life-time sexual partners was 1.67. Of the women in relationships, the partner was the father of the child in all but one instance. On average, the women's partners were 5.29 years older than the women (average age=24.3).

Housing Conditions

ALL OF THE women lived in brick houses, with an average of 3.4 rooms (not including the bathroom). Thirty-eight percent (n=7) lived with their partners; sixty-one percent (n=11) lived with their own parents. On average, in a house, 2.4 people lived in each room. Houses themselves were made from brick, usually with a concrete floor. They all had electricity and a television. Although some homes were more modest, with dirt floors and fewer amenities, they were generally similar to the living conditions for the majority of the population in Praia Grande.

Municipal and state-provided services in Praia Grande contributed to a less than ideal health environment. Houses had a connection to the municipal water source, but this connection functioned sporadically. Every month, a representative of the city municipal health system came to treat with chlorine the rain/well barrels kept as back-up. This treatment helped to avoid mosquitoes that laid eggs in the barrels and that might spread dengue. Unfortunately, an excessive

amount of chlorine often turned the water thick and white. As a result, some families hid the barrels from the city officials.

The sewage system of the houses involved the construction of a *fossa*, or concrete-lined hole, into which the feces and urine traveled. When the *fossa* filled or the day was hot, the sewage often bubbled up onto the streets to flow down a shallow channel in the middle of the narrow road before dumping into the bay and onto the mudflats where the women gathered shellfish.

Income and Occupation

THE AVERAGE FAMILY income did not reach Brazil's minimum wage, with an average of R\$395 per month (US\$244) for a family of average 6.3 persons. Minimum wage is currently R\$540 per month (US\$340 dollars) for a family of four (Governo Federal do Brasil). Less than R\$300 per month for four persons is considered "indigent" by the Brazilian government (Rocha, 2008). Half of the women who participated fell into this category (n=9). In all but two cases, where the woman's mother or mother-in-law was considered the primary source of family income, the primary wage earner was either the woman's father (50%, n=9) or partner (39%, n=7). In the two households headed by women, the family incomes were R\$200 and R\$300 per month.

According to the CHW, the majority of families collect checks from the Bolsa Família program, which provides money to families with dependent children as long as those children remain in school. With support from this program and the sustenance from the shellfish, fish, and fruit of the island, hunger did not seem to be common, and the CHW noted that malnourished children were rare.

Education and Occupation

ON AVERAGE, THE women studied in a school for 7.8 years. All reported periodic interruptions in their schooling before the pregnancy. Two of the women (11%) graduated from high school; sixteen women (89%) reported leaving school because of the pregnancy or to take care of the child. On average, their partners spent 7.6 years in school.

Eleven of the women, more than half, did not consider themselves to have a profession. Of those who reported one, four (57%) said they were *marisqueiras* or shellfishing women. Of the other three, one woman reported her occupation as mother, one as an artisan, and the last as a facilitator in a local activist organization.

The majority of the husbands and partners had professions involved with construction (22%, n=4) or fishing (39%, n=6). In general, construction jobs had the highest average income because many of them were state jobs requiring that the men earn minimum wage. The rest of the women described their partners' or fathers' professions in manual labor or odd jobs with a variety of terms that had similar meanings, such as *aventureiro* (adventurer), *não tem (profissão)* (no profession), *autônomo* (autonomous), or *serviços gerais* (general services). Most men on the island were involved in the fishing industry in some way and participated in the basket making. Older men, from the ages of 30 to around 60, generally took the baskets to the market to sell on the weekends.

Reflections on Motherhood and Changes in Social Relationships

AFTER BECOMING PREGNANT, the primary effects the women reported were the changes in their relationships with their friends, parents, and partners. While their time with friends was dramatically decreased, motherhood brought them closer to their own mothers and generally strengthened their relationships with their partners. The community's disapproval of unfaithful fathers seemed to play a key role in the transformation of the relationship. The pregnancy also began the process of a couple's informal marriage, symbolized through the construction of a home. Given the mixture of the loss of friends and freedom with the gains of expanding their family networks, the young women's overall ambivalence towards their pregnancies and motherhood becomes understandable.

The majority of pregnancies appeared to be neither planned nor entirely unplanned. In most cases (12 out of 18), the young women were taking hormonal birth control or were using condoms with their partners before the pregnancy. Six of the women were not using a form of pregnancy prevention, although they did not explicitly plan to become pregnant. Three of the women said their pregnancies were due to birth control failure, whereas two of the women cited a lack of knowledge about or access to birth control as the most important factor. Therefore, only 27.8% (n=5) of the pregnancies of the young women can be attributed to traditional public health targets for preventing *unwanted* pregnancies: lack of access to proper birth control use and information on how to use it.

Of the women who were personally interviewed in depth (n=5), both those who were using birth control (n=3) and those who were not using birth control (n=2) when they became pregnant expressed

ambivalence about becoming pregnant, saying that they remembered their own vague desires to become a mother or recounting their boyfriends' desires to become a father. Combined with survey data that more than two-thirds of the women did not consider their pregnancies to be a mechanical or informational failure, this initial data suggests that the pregnancies, although not explicitly planned, were not *entirely* undesired in most cases.

Just as the series of events that led up to the pregnancy varied from a moment of passion and intoxication during a Carnival festival to the celebration of the couple's newly constructed home, the emotional reactions of the women to learning about their pregnancies were highly individual. The following table shows the five women's reactions to their pregnancies:

Subject	Portuguese	English
1	<i>Fiquei triste no começo, mas minha mãe e namorado gostavam tanto que acabei gostando...ao fim, a vida não é muito diferente, só como uma responsabilidade a mais</i>	I was sad at first, but my mom and boyfriend liked it so much that I ended up liking it too...in the end, [being a mother] isn't that different, just one more responsibility.
2	<i>Senti-me chocada. Sabia que a gente não tava usando camisinha diretinho e tudo mas não achei que ia ficar grávida.</i>	I was shocked. I knew we hadn't been using the condom right and everything but I didn't think I was going to get pregnant.
3	<i>Fiquei muito feliz porque fazia tempo que minha mãe tava querendo...ela só tinha menina e nunca menino e queria sempre um menino para criar...Fiz minha mãe muito feliz.</i>	I was really happy because my mom had been wanting it to happen for a long time...she only had girls, never a boy and always wanted to raise a boy...I made my mom really happy [when she became pregnant with a boy].
4	<i>Não teria escolhido, sabe, que senti saudade de ir aprontar com minhas amigas. Tive que ficar uma mulher séria.</i>	I wouldn't have chosen it, you know. I missed going out with my friends and getting into trouble. I had to turn into a serious woman.
5	<i>Chorei e chorei mas agora a emoção passou...adoro minha filha...Era difícil porque o pai era longe e não ligou para mim.</i>	I cried and cried but now that passed...I adore my daughter. It was hard because the dad was far away and didn't care about me.

The group discussion also emphasized the variety of reactions the women had to their pregnancies:

I drank a strong tea to regularize my menstruation but it didn't come.

I (emphasis) didn't take any medication. (in response to the other mother)

My sister worked in a pharmacy and offered to bring me something for me. I talked to my mom, and she said that it was my choice if I wanted to have an abortion...I got scared that [the child] would come out defective, and I didn't go.

...I tried to get rid [of the child] with a strong tea but it didn't work...she was too strong and so that was that. [The father's] aunt got really mad at me when she found out and tried to take my daughter away from me. She said I wasn't old enough to understand and take care [of a child].

[CHW] Oh my child, you can't do that...

I cried, cried, cried, cried, but I never thought about getting rid of the child.

My aunt wanted me to get rid of the baby, but I got pregnant and I had to take care of what I made.

I didn't want to get pregnant, but it is what God wants...

In their personal and group narratives, the women related their initial personal reactions as secondary to the eventual acceptance and support they received from the father of the child and their mothers. Though the pregnancy came as a primarily negative personal experience, it was then re-imagined as a positive, or inevitable, experience that was largely supported by their loved ones.

Out of the larger group of nineteen, thirteen of the women reported that the pregnancy primarily affected their social lives. The pregnancy forced the young women to leave their peer groups and suddenly move into womanhood. For many, the transition was interpreted as the end of their personal lives (when they would *curtir*, *beijar*, *namorar* or have fun, kiss, make love). Some young mothers resented the change. Two of the interviewed women regretted not having finished high school (*nem me formeí da escola*). During the group session, three of the women commented:

If I could, I wouldn't be a mom. Now I have two kids and I don't have any more friends. But what are you going to do?

Me too. My own life is over.

I like being a mom, but it's just a lot of work. Being a mom and a housewife is really hard. It makes me nervous to take my kid to that school [on the mainland, where many families

reported their children being beaten by gang members who did not like the *quilombolas*].

Personal interviews also revealed different levels of ambivalence and mixed feelings about motherhood:

Being a mom is good when it's planned. Without planning, it's hard.

It's not that I wish I hadn't had my son, but if I could tell others, I would tell them to have their own lives first.

Despite this recognition of the difficulties of motherhood and the mourning of the loss of their own childhoods, the young women also attributed much of what was good in their lives to their children. The pregnancy and motherhood constructed different social networks, usually strengthening the same ones that they would have as adult women in the community. All but one of the women, who was estranged from her mother prior to the pregnancy, reported that their relationships with their mothers improved after the birth of the child. Many had a new respect of their mothers whom they had dismissed in the past.

[Group meeting] I was the youngest and didn't know how to take care of a kid...so she [her mother] showed me...it's like I gained a new respect for my mom. One son is already too much for me, and she raised six!

[Personal interview] When my son was little, I always wanted to take him to shellfish with me. And my mom fought with me. She said that he was going to burn in the sun, and I couldn't take him. I fought with her; I don't know why... we always fought just to fight. But I thank her now because my son is already two and he might not be white like you [pointing to me] but he isn't black like me. So I thought, my mom really does know something.

[Personal interview] My husband fell in love with our son. And now we are in our house, and he has work, thank God, and doesn't do that stuff anymore [such as the beatings discussed earlier]. Our son is our happiness.

Despite the rarity of the study participants mentioning violence during the research, many highlighted the stress caused by their partners' infidelity prior to the birth (however, this infidelity disappeared after their partners became fathers). Based on the CHW's dismissive remarks to one mother that her separation from her boyfriend because of his infidelity was only temporary, there seemed to

be a strong social expectation that the parents of the child would become an official couple. This impression was reinforced by casual conversations, in which multiple people cited a particular cheating father and everyone's reprimands for him to stop his infidelity for the sake of his daughter. Perhaps as a result, almost all participants who mentioned infidelity said that after the birth, their previously cheating boyfriends transformed into faithful husbands and fathers, though sometimes the transition was far from instantaneous.

During the life history, the young mother emphasized on multiple occasions the planning of her house with her husband as strong symbolic proof of their union and the end of his previous cheating behavior.

I decided that I was going to take control of my life. So I went to the church and asked God to change him. And since we have lived together, I can feel the difference. He started to buy things for our daughter. He even bought these little sandals for her. And last week, he even took me to a party. He never took me to parties, because he was ashamed of me... [I tell him every day] 'you should kiss my feet for what I suffered. Because only God knows how much I suffered. But I have won now, and now I am in my house.'

Because the construction of the house required a great deal of mutual work and planning, it acted as a concrete symbol of a couple's permanent union. Most of the mothers reported that they chose not to "marry on paper" or hold a ceremony because that required money that they wanted to put towards their houses. According to the CHW, in Praia Grande, the shared home had become a strong symbol of social relationships and bonds, more important than official marriage or ceremonies.

Overall, the young women's experiences highlighted the ambivalence they felt towards pregnancy and motherhood. Many mourned the loss of their friends and schooling opportunities; yet, through the pregnancy, the women often strengthened ties with their own mothers, who helped them to raise the child, as well as with the families of their partners. The child's birth also marked the positive beginning of the couple's planning to construct their own house and space on the island, representing their bond and long-term future together.

Infrastructure and Dream

THE INADEQUACIES OF infrastructure generated a lively discussion among the young women, which began when the one mother

who was not born on the island said, “*Não vou mentir; gosto mais de Salvador. Aqui não tem nada.*”/I am not going to lie, I like Salvador more. Here there isn’t anything.” Other women voiced their opinions of what the island did not have. In the ensuing clamor, the majority agreed that what Praia Grande most needed was more paid jobs for their husbands.

Another primary complaint was the lack of police. In the prior six months, two murders had occurred on the island, generating a great deal of concern and fear because no murders had occurred in decades. Because of the lack of police, there did not appear to have been a formal investigation. Despite the low level of violence on the island itself until recently, most of the women had lost brothers or cousins in Salvador. As a result, many avoided going to the mainland for fear of being assaulted or robbed on buses. Two of the women had already been robbed in this manner.

In addition to the primary concerns over lack of police and jobs, the young women declared the need for a twenty-four-hour doctor living on the island, schools, a grocery store, retail stores, and a place to get an identity card. Based on the responses of the young women and their desires for biomedicine, police, and shops, the age of commercialization in Praia Grande seems imminent. While the dreams and goals the women reported may reflect their interpretations of what they thought I (an American researcher) would want to hear, their recognition of what an outsider would consider legitimate goals of education and wage work also reflects the rapidly increasing interface with Brazilian society and the world:

My dream, I want to graduate. I want to teach children.

Mine too.

Mine is to graduate middle school.

I want to practice law.

Man, I want to be someone in life. From here on out.

To be employed.

I want to be a doctor (the rest of the women laughed).

The ways in which biomedicine and other post-industrial constructs will play out on the island largely remains to be seen. From the continuation of the use of contraceptive teas to the binding of pregnant women to support the baby in later months of pregnancy (both discussed below), many *quilombo* traditions still exist, and not all of the

young women have fully taken up the “modern” dreams for their own employments. During the group meeting, while their peers expressed dreams for employment, two women still responded:

My dream is for my husband to work.

Oh my god, my dream is...I don't know.

Significantly, both in the group session and in the personal interviews, the question about personal goals led to laughter and to delayed answers. Most resisted answering or copied their peers' answers. Many knew the socially correct answers, but few during the personal interviews were able to express that they might achieve the goals or what they might do.

When the young mothers were asked both in the personal interviews and group meeting what opportunities they felt they had missed as a result of their early pregnancies, their universal response was silence. Whether a mother or not, young women will not likely become doctors or lawyers after attending the public schools on the island.

Due to the rough sea, preventing access to school many days, and the poor pay of the teachers, children from Praia Grande generally miss about a third of the days of school. Combined with the effects of bullying from gangs and the generally poor quality of teaching (at least as reported in informal interviews by the parents of the children), the chances of Praia Grande children achieving a college education are even less than the chances of other Brazilian public school children. Based on the interviews and conversations, only one island child was able to pass into a university, and he did not live on the island after the age of ten.

Interaction with Biomedicine

MOST OF THE young women held biomedical knowledge in reverence and desired more medical professionals on the island. A doctor had started to come to the island about six months before the beginning of this research for about five hours per week. However, many of the women were still frustrated by their inability to get a pediatric appointment for their children. They waited hours in the clinic for an appointment. By the time the physician arrived at 10:00 or 11:00 o'clock in the morning, there was often a line of thirty people waiting outside the clinic; only six or seven would be seen. After waiting in the clinic on several occasions without being seen, one mother had to take her son to the hospital on the mainland, where she spent the

whole day waiting for him to receive nebulization.

Despite the violations of their rights to health care under the Brazilian Constitution and their frustrations with specific physicians, the young women were unwavering in the prestige that they attributed to a doctor's opinion. In fact, the guise of medical expertise was used to defend the women's own understandings, as seen in the following exchange between the CHW and one young mother:

I am using a [birth control] injection without hormone because I got really, really fat with the other one.

CHW: All [birth control] injections have hormone, my daughter (affectionate term).

This one doesn't. The doctor said so.

The clinic's construction increased access to Depo-Provera and the percentage of women who used it as their birth control. According to the health clinic nurse, she and the new doctor generally promoted it for women for whom the oral birth control had "not worked" the first time. In a stable relationship, condom use was not expected or considered reliable. As one of the young mothers noted, "*Não usamos camisinhas porque não funcionam bem. Melhor injeção.*" We don't use condoms because they don't work. [Birth control] injections are better."

The suggestion of injection had been internalized by many women as a fail-safe way to prevent a second pregnancy or a convenient way to not forget pills. Seventy-two percent (n=13) of the young mothers were using an injection birth control during the time of the study (although three also used condoms sometimes), accounting for 94% of the total birth control use.

The percentage is largely reflective of general community statistics. According to my count of the clinic records, 74% of the 115 women who had received birth control from the clinic in May 2010 received a three-month hormonal injection. Nonetheless, one of the twice-pregnant mothers had been taking Depo-Provera and "forgot" an appointment. In private, she admitted that her husband wanted their child to have a brother or sister. None of the young mothers who reported using the hormone injection thought that it would have a negative impact on their health in the long-term; the primary problem reported was weight gain.

The young mothers' kinship ties allowed a dialogue and practice that often diverged from the medical professionals' instructions. Despite the high esteem in which they held doctors, young mothers often ignored

the doctors' suggestions when they conflicted with community knowledge. For example, one of the women professed that the oral birth control pills made her become pregnant. Afterwards, she returned to using the strong teas made by her mother. Three of the women in the personal interviews also reported that they did not believe the doctor's claim that one pill a day was sufficient. In addition to their injections, they bought oral contraceptives from the mainland and would take a few pills after unprotected intercourse. In many instances, the young women integrated traditional advice from older sisters or the mothers about how to prevent pregnancy with the new contraceptive medicines.

Despite the increasing medicalization of reproduction, apparent in the Depo-Provera prescription increase that is occurring on the Ilha de Maré, young mothers and women should not be regarded exclusively as passive victims of social processes. They are active agents who interpret the data and information handed down to them by authority figures, reshaping that knowledge to coincide with their world paradigm. Nonetheless, it is important to remember such re-interpretation occurs within an extremely unequal difference in power between the medical authorities and the young mothers.

Given the generally negative experience of most Praia Grande residents with the official Brazilian state institutions, the young mothers' continued attraction to them seems both odd and optimistic. Most of the interviewed women could relate at least one instance in which they had felt racism in a clinic, hospital, or school setting. However, they continued to express a desire for full citizenship and their rights for education and healthcare. If further ethnographic studies indicate that these responses are not simply a reflection of their identifications of me with the social institutions, this positive attitude towards state institutions would promote a more thorough integration of more extensive health care services into the community.

DISCUSSION

THE PARTICIPANT OBSERVATION I conducted on the island points to the structural violence experienced by the *quilombo* population and the need to assess cultural contexts of adolescent or young motherhood before implementing reproductive health programs. Ultimately, through the disciplinary lenses of anthropology and public health, this research challenges the notion that teenage pregnancy is an undesirable health outcome in all situations and has a negative impact on education and socioeconomic status.

The narrations of the women reveal an active dialogue occurring in Praia Grande with the larger social discourse about what makes a woman valuable. For example, in the life history, a relatively rich mother-in-law considered the young mother's lack of education to be a social disqualification for marriage or partnership with her son. Likewise, the young mothers were well aware of the sorts of goals and dreams that white women from outside the island would consider legitimate.

Without dismissing the relevance of the young mothers' dreams of college education, it is worth noting that no one with whom I spoke on the island could think of someone raised on the island who had been able to enter a university. One son of an island mother had been adopted early on by relatives in Salvador and had managed to enter UFBA; however, while the people of Praia Grande were proud of him, his educational opportunities were clearly different than those of the people raised in the quilombo. The professional futures of the young women in Praia Grande were defined and limited much more by social structures and the inadequacy of state institutions than by their early pregnancies.

In addition to the failings of the school system, racism deeply affected the opportunities of the young women. While the people of Salvador were mostly ignorant of the existence of the *quilombolas*, the *quilombola* women were very aware of their own places in the Brazilian social system. Given the violent mistreatment they received from people not from the island, the reluctance of the young people to leave the island to go to high school and pre-vestibular classes (yearlong courses that train young people to take the college entrance exam) is more than understandable. Even the community health worker felt unable to discourage her daughter from quitting the pre-vestibular course in the city (after the near constant abuse and discrimination that her daughter experienced in the city from being both black and rural) to return to Praia Grande to begin constructing her house with her husband.

One woman commented, "The reason we like [the SIT study abroad director] Damiana's students here [in Praia Grande] is because you treat us like humans. We aren't animals to you. The white Brazilians, they come here but they would never sit down and eat with us. You sleep in our homes with no problem. You eat our food. That is why we always welcome you with open arms."

My racial identity and foreign origins probably drew out responses from the study participants based on their perceptions of

what values would be held by a young, white woman from the U.S. without any children. The research by Jovânia de Silva, one mentor of this project, showed the depth and breadth of the effect of my white, American identity on what version of the truth the mothers communicated. A graduate student and Afro-Brazilian, Jovânia was also a mother. She found that for the adolescent women, “*a gravidez apesar de não ter sido planejada, é percebida como prazerosa*”/Despite not being planned, the pregnancy is often perceived as pleasurable.” While the women tended to highlight the shock and sadness of the pregnancy to me (a childless, white woman), especially in the earliest interviews, the pleasurable aspects were more acceptable to share with another mother presumed to better understand the desirability and joys of children. Only in the group session, talking among themselves, did the second theme emerge in my own research. However, it was precisely this difference in reported experience that shows that a double-consciousness of social ideals and realistic goals has begun to be established on the island.

By the end of my time on the island, the women’s responses began to be more nuanced about their pregnancies and motherhood. Women who initially had stated that the pregnancy ended their lives began to describe their children as the source of their happiness and as the basis of the legitimacy of their relationships with the child’s father. The pregnancy and motherhood represented an important connection back to their own community, integrating them into the complex kinship network of Praia Grande in a way that education could not. Girls were transformed into women who inherited the generational knowledge of childcare from their mothers, initiating their new social role. In the case of one of the participants, the rumors of her alcoholism and constant fighting led her natal family to disown her. However, her social self was reconstituted upon the birth of her first child when her mother-in-law’s family adopted her and supported her child.

Early pregnancy can be expressed in the practicality of strengthening social relationships. The strong kinship ties in the community were the best guarantee of social support during times of hardship. Through pregnancy, a young woman gained an additional household (her mother-in-law’s) on which she could count to support her during times of hardship. Mother-in-laws’ households and extended family were essential sources of food and monetary support during difficult times. By linking herself into a new social network through pregnancy, a woman gained not only a new social role but also insur-

ance against economic hardship. During the middle of one interview, the young mother's nephew brought in some *feijoada* (beans, rice, meat) to feed the new family for the day, since the husband's fishing accident left the new family in dire straits. She then said, "without my daughter, I would still be in my aunt's house, eating the scraps they didn't want. Now, I have my house, and we eat *feijoada*."

It is not the intention of this research to give the impression that the mothering experiences of the young women were uncomplicated. Some of the young women had much more negative experiences with motherhood, especially those who failed to develop the familial support network following the pregnancy. For example, one young mother had encountered many problems with her mother-in-law, who had other women in mind for a prospective wife of her son.

Domestic violence was also an issue. In a personal interview, one woman related how her husband beat her up while pregnant and brandished a knife against his mother. Although the couple had bettered their relationship by the time they were interviewed, it is unclear how long the resolution would last or if other problems would develop. Despite the suffering the women experienced as a partial result of their social responsibilities as mothers, it remains unclear if the problems in their relationships are unique to their age or if they reflect more broad-spread patterns of difficulties that Praia Grande mothers face.

Finally, the most widespread negative result of the pregnancy for the young women seemed to be their leaving school. Although, as previously noted, this probably did not affect their career choices or future earning potential, education could be seen to have its own intrinsic merits. Leaving school also seemed to decrease the women's peer networks, which could be especially significant in cases of domestic violence. However, most of the women reported that they would not have left school because of the pregnancy if there had been options on the island. Therefore, the negative impact of the loss of schooling seems to be as much a result of inadequate infrastructure as the age of pregnancy.

With the aforementioned exceptions clearly in mind, in general, young women in Praia Grande use motherhood to construct an important piece of their social selves. The support the young women gain from their own mothers and, often, partners results in a transformation to acceptance and joy of producing a new life. Moreover, the pregnancy gives these women a social claim over

their partners and begins the path to marriage and establishment of their households.

CONCLUSIONS AND FUTURE RESEARCH

ALTHOUGH THE WOMEN who participated in this study generally fit the epidemiological profile of adolescents who become pregnant in the United States and Brazil — poor, less educated, and a racial minority — their experiences of pregnancy and motherhood differed significantly from woman to woman. While some regretted the pregnancy, most had taken to the role of mother with apparent joy and familial support.

Moreover, the critical question included in the interviews, with the idea of demonstrating that further access to free birth control and information would increase women's education and job prospects, failed to illicit the response I expected. None of the young women could respond as to what, if any, opportunities they had missed as a result of their pregnancies. Their answers led me to question my own basic assumptions about family planning.

In Praia Grande, I found that island poverty was not perpetuated by the age at which women chose to have children but instead by the dearth of other options for advancement through the market economy and educational system. Moreover, the patterns of shared child rearing I witnessed on the island call into question the assumption that the adolescent mother should or will accept sole responsibility for her child's development. With a broader kinship network in place, the concerns about young women's abilities to support and provide adequate care for their children became largely irrelevant.

Kinship ties gained from motherhood and conjugal unions hold both emotional and practical implications, especially in this environment of poverty in which social support acts as one of the primary buffers to extreme hardship during times of individual failure or misfortune. However, the increasing contact with the outside through television and other media has created a consciousness that these forms of kinship and early pregnancy are not viewed favorably within the dominant culture. Such a perception introduces an element of shame for not having completed basic education, even when that may be a reasonable choice given a woman's circumstances.

This argument is far from a novel one. As seen in the literature review, a number of researchers from public health, anthropology,

and sociology have come to the same conclusions. Despite this, prevention of adolescent pregnancy has become a prominent public health agenda for organizations like the CDC, WHO, and SUS. All three of these organizations base their agendas on the premise that adolescent pregnancy contributes to the continuation of poverty and poor health outcomes for young women. That premise is not supported by my primary research and is a reductionistic treatment of a complex socioeconomic phenomenon.

This research is far from complete. Many gaps remain in my own understanding of the *quilombo*. Three months of research, including only one in which I lived in the community, were hardly sufficient to understand the complexities of motherhood and sexual health in Praia Grande. I hope to fill some of these gaps when I return for a year of field research in 2012 through a Fulbright grant.

The provisional nature of the conclusions conveyed here reinforce the principle point that close, long-term, and interpretative research is valuable and necessary in order to evaluate preconceived assumptions held by outsiders, including academics and professionals who intend to help. Because such assumptions are based on the professional's own background and experience, they may be erroneous outside of her or his own context. Such was the case in my own time on the island, where my basic notions of family planning, education, and life prospects were often not applicable or relevant.

Based on the results of this study, I propose that an interdisciplinary theoretical framework can use ethnographic methods to ground public health projects in their specific geographic, historical, and cultural contexts. The application of the theoretical framework in Praia Grande suggests how medical professionals and public health practitioners might evaluate the cultural bias implicit in their reproductive health programs' agendas. As Geronimus (2003) says:

Cultural dominance can be perpetuated by well-meaning people consciously dedicated to children's well-being, social justice, and the public good. The entrenched cultural interdependence of and social inequality between European and African Americans leads African Americans to be highly visible targets of moral condemnation for their fertility behavior, and also sets up African Americans to pay a particularly high political, economic, psychosocial, and health price. (p. 881)

Without provision of education and meaningful occupational opportunities, *quilombo* adolescent women, like poor black women in the U.S., are likely to pay a high price for modernization and industrialization without reaping the potential benefits. Though it is necessary to question the efficacy of biomedicine and modern ideals, the processes of globalization are unlikely to slow down because of the ways that social mores benefit individuals in power. Moreover, the long-term and concerted community efforts to establish the clinic and schools show Praia Grande's residents' interest in development; therefore, attempts to slow their implementation are paternalistic and disrespectful of the community and its leaders' rights to determine what is best for the *quilombo*. According to the community health worker, the construction of the clinic resulted from a nearly twenty-year-long struggle with the municipal government to provide the funds.

The effect of close ethnography in this case should not be for the well-meaning public health officials, medical professionals, and social scientist researchers to shrink away from the enormous challenge of providing reproductive health services while recognizing their own ideological prejudices. Rather, close, interpretative research can help to shape professionals' awareness of their constructed and privileged subjectivities, therefore helping to guide better public health practices that reflect the desires and subjective well-being of the women for whom the services are designed.

The complexity of the women's experiences as *quilombo* residents and as young mothers can be overwhelming. Despite the difficulty of interpreting the often ambiguous and contradictory responses collected through this fieldwork, one clear idea emerged: pregnancy for these young women was not the end of a life trajectory as adults and community members but instead the beginning. As such, their stories reveal more about the injustices, racism, and difficulties that Praia Grande and *quilombo* women face than about the consequences of becoming pregnant while young.

APPENDIX: INFORMED CONSENT AND SURVEY

Consentimento da participação no estudo “Mães adolescentes numa comunidade quilombola, Ilha de Maré”

Nesse dia _____ de 2010, eu _____

consinto em participar no estudo com o título acima, entendendo que é ligado com o grupo de pesquisa CRESCER dentro da Escola de Enfermagem de UFBA e também com o programa de intercâmbio SIT (School for International Training). Entendo que embora já tivesse consentido, posso mudar meu mente em qualquer momento e não preciso responder a uma pergunta específica se não me sinto cômoda. Entendo que meu nome não vai ser publicado em qualquer revista ou outro documento pública e informações que eu dê não serão publicadas numa maneira que deixaria alguém me identificar.

Assinatura _____

[translation]

Consent for participation in the study: “Adolescent mothers in a community quilombola: Ilha de Maré”

On this day _____, 2010, I _____

consent to participate in the study with the title above, understanding that it is connected to the group CRESCER in the Nursing School of UFBA and also with the Exchange program SIT (School for International Training). I understand that even though I have already consented, I can change my mind at any time and I do not need to respond to any specific question if I do not feel comfortable. I understand that my name will not be published in any journal or other public document and that information that I give will not be published in a way that will allow someone to identify me.

Signature _____

QUESTIONÁRIO PRELIMINAR

*Perguntas sem resposta serão marcadas 99; se não aplicar, serão deixados sem preencher

CÓDIGO DE REFERÊNCIA:

Nome completo _____

Dato de nascimento _____

Idade _____

Profissão _____

Estado marital _____

O que você faz atualmente? 1-Estuda 2-Trabalha 3-Cuida filho

Com quem mora: 1-Sozinha 2- Pais 3-Parceiro 4-Outro

Há quanto tempo que mora nessa casa: _____

Quantos lugares no último ano: _____

Quantas pessoas moram na casa: _____

Qual a renda familiar: _____

Quem ganha a maioria da renda familiar: _____

Quanto ganha você: _____

Parceiro tem: 1-Sim _____ (idade) 0-Não

Pai do seu filho: 1-Sim 0-Não

Profissão dele: _____

Nível de escolaridade: _____

Quanto ganha: _____

Tem carteira profissional assinada: 1-Sim 0-Não

Número de filhos: 0 1 2 3 4

Idade dos filhos: _____

Qual o seu nível de escolaridade: _____

Razão para deixar: _____

Quantas gestações: 0 1 2 3 4

Número de partos: 0 1 2 3 4

Já teve aborto: 1-Sim _____ (espontânea ou não) 0-Não

Número de partos cesários: 0 1 2 3 4

Idade da primeira gravidez: _____

Idade da primeira menarca: _____

Idade da primeira gravidez da mãe: _____

Na última gravidez, consultas feitas ao pré-natal:

Onde _____

Número de consultas: 0 1 2 3 4 5 6 7 8 9 10+

Teve algum problema de saúde durante a última gestação?

1-Si 0-Não

Planejamento familiar atualmente:

0-Não

1-Preservativo/Camisinha masculino

2- Preservativo/Camisinha feminino

3-Espermicidas

4- Diafragma

5- Anticoncepcionais orais

6- Anticoncepcionais injetáveis

7- Implantes

8- Dispositivo intra-uterino

9- Amenorréia da lactação (LAM)

10- Métodos cirúrgicos

11-Percepção da fertilidade (“natural”): muco-cervical, temperatura basal, método sintotérmico, calendário

Número de parceiros: 0 1-3 4-6 7-10 10+

SITUAÇÃO DA MORADIA E SANEAMENTO

PARTICIPA DE GRUPOS COMUNITÁRIOS	
Cooperativa	
Grupo religioso	
Associações	
Outros-Especificar	
MEIOS DE TRANSPORTE QUE MAIS UTILIZA	
Ônibus	
Caminhão	
Carro	
Carroça	
Outros-Especificar	

EM CASO DE DOENÇA PROCURA	
Hospital	
Unidade de Saúde	
Benzedeira	
Farmácia	
Outros-Especificar	
MEIOS DE COMUNICAÇÃO QUE MAIS UTILIZA	
Rádio	
Televisão	
Outros-Especificar	

TRATAMENTO DA ÁGUA NO DOMICÍLIO	
Filtração	
Fervura	
Cloração	
Sem tratamento	
*Frequência de problemas	
ABASTAMENTO DE ÁGUA	
Rede pública	
Poço ou nascente	
Outros	
DESTINO DE FEZES E URINA	
Sistema de esgoto (rede geral)	
Fossa	
Céu aberto	

TIPO DE CASA	
Tijolo- Adobe	
Taipa revestida	
Taipa não revestida	
Madeira	
Material aproveitado	
Outro- Especificar:	
Número de Cômodos /Peças	
Energia Elétrica	
DESTINO DO LIXO	
Coletado	
Queimado-enterrado	
Céu aberto	

OUTRAS INFORMAÇÕES

Alguém na família possui plano de Saúde? _____

Número de pessoas cobertas por Plano de Saúde? _____

Nome do Plano de Saúde: _____

OBSERVAÇÕES (casa, alimentação, criança, tempo, relações)

PRELIMINARY QUESTIONNAIRE

*Questions without a response will be marked 99; if it does not apply, leave blank without being filled out.

DE-IDENTIFIED CODE:

Complete name _____

Date of birth _____
Age _____
Profession _____
Marital status _____

What do you do right now? 1-Study 2-Work 3-Take care of child

With whom do you live? 1-Alone 2- Parents 3-Partner 4-Other

How long have you lived in this house: _____

How many places in the last year: _____

How many people live here: _____

What's the family income: _____

Who makes the majority of the income: _____

How much of it do you make: _____

Your partner: 1-Yes _____ (age) 0-No

Father of your child: 1-Yes 0-No

Profession: _____

Education: _____

How much does he make: _____

Do you have a <carta assignada>: 1-Yes 0-No

[Translation note: Carta assignada is given when an individual is officially employed and given benefits like social security and worker's compensation during periods of injury.]

Number of children: 0 1 2 3 4

Age of children: _____

Highest grade completed: _____

Reason for leaving school: _____

Number of pregnancies: 0 1 2 3 4

Number of births: 0 1 2 3 4

Have you had an abortion/miscarriage: 1-Yes _____ (which) 0-No

Number of Cesarean sections: 0 1 2 3 4

Age of first pregnancy: _____

Age of first period: _____

Age of mother's first pregnancy: _____

In the last pregnancy, did you complete prenatal care:

Where _____

Visits: 0 1 2 3 4 5 6 7 8 9 10 10+

Did you have a health problem during the last gestation?

1-Yes _____ 0-No

Actual family planning strategy:

0 - None

1- Male condom

- 2- Female condom
 - 3- Spermicide
 - 4- Diaphragm
 - 5- Oral contraceptives
 - 6- Injectable hormonal
 - 7- Implants
 - 8- IUD
 - 9- Lactation (LAM)
 - 10- Surgery
 - 11- "Natural:" Perception of fertility, mucocervical, basal temperature, "pulling out," calendar
- Number of lifetime partners: 0 1-3 4-6 7-10 10+

SITUATION OF LIVING AND SANITATION

[Translation note: This part of the survey is taken from the Sistema Único de Saúde's (national health system's) national health survey.]

PARTICIPATION IN COMMUNITY GROUPS	
Cooperatives	
Religious groups	
Associations	
Other: specify	
METHOD OF TRANSPORT MOST USED	
Bus	
Truck	
Car	
Tractor	
Other: specify	

IN CASE OF SICKNESS, LOOK FOR	
Hospital	
Health post	
Benedeira [herbalist/prayer]	
Pharmacy	
Others: specify	
METHODS OF COMMUNICATION MOST USED	
Radio	
Television	
Others: specify	

WATER TREATMENT IN HOME	
Filtration	
Boiling	
Chlorination	
No treatment	
*Frequency of problems	
WATER SOURCE	
Municipal	
Well or spring	
Other	
DESTINY FECES AND URINE	
Sewage (public system)	
Hole	
Open sky	

TYPE OF HOUSE	
Brick/adobe	
Siding (treated)	
Not treated siding (e.g. uncoated aluminum)	
Wood	
Loose material	
Other: specify	
Number of rooms	
Electric energy?	
DESTINATION OF TRASH	
Collected	
Burned/buried	
Open sky	

OTHER INFORMATION

Does someone in the family have a [private] health care plan? _____

Number of people covered by the health care plan _____

Name of the plan:

OBSERVATIONS (house, feeding/food, child, time, relationships)

REFERENCES

- Adaszko, A. (2005). Perspectivas Socio-antropológicas sobre La adolescência, La juventud, y el Embarazo. *Embarazo y Maternidad en la Adolescencia. Estereotipos, Evidencias y Propuestas para Políticas Públicas*. Buenos Aires: CEDES-UNICEF, 33-65.
- Araújo dos Anjos, R. (2000). A Questão dos Remanescentes de Quilombos no Território Brasileiro. *Observatorio Geografico De America Latina*, 1-5.
- Barker, S. L., & Castro, D. M. F. (2002). Gravidez na adolescência: dando sentido ao acontecimento. In M. L. J. Contini, S. H. Koller & M. N. S. Barros (Orgs.), *Adolescência & psicologia: concepções, práticas e reflexões críticas* (pp.78-84) Brasília: CFP.
- Barlow, K., & Chapin, B. (2010). Special Issue: Mothering as Everyday Practice. *Ethos*, 38(4), 321-460.
- Berenson, A. B., Wiemann, C. M., & McCombs, S. L. (1997). Adverse perinatal outcomes in young adolescents. *Journal of Reproductive Medicine*, 42, 559-563.
- Bukulmez, O., & Deren, O. (2000). Perinatal outcome in adolescent pregnancies: a case-control study from a Turkish University Hospital. *European Journal of Obstetric and Gynecological Reproductive Biology*, 88, 207-212.
- Caldas, A., & Garcia, L. (2007, August 13). Direito a Terra das Comunidades Remanescentes de Quilombos: o longo e tortuoso caminho da titulação. *Justiça Global*. Retrieved May 12, 2011, from <http://global.org.br/programas/direito-a-terra-das-comunidades-remanescentes-de-quilombos-o-longo-e-tortuoso-caminho-da-titulacao/>
- Card, J. J., & Wise, L. L. (1978). Teenage mothers and teenage fathers: The impact of early childbearing on the parents' personal and professional lives. *Family Planning Perspectives*, 10(4), 199-205.
- Conde-Agudelo, A., Belizan, J., & Lammers, C. (2005). Maternal-perinatal morbidity and mortality associated with adolescent pregnancy in Latin America: Cross-sectional study. *American Journal of Obstetrics and Gynecology*, 192(2), 342-349. doi: 10.1016/j.ajog.2004.10.593
- Conrad, R. E. (1985). *O Tráfico de Escravos para o Brasil*. São Paulo, SP: Brasiliense.
- Costa, M. O., Santos, C. T., N. Sobrinho, C. L., Freitas, J. O., Ferreira, K. L., Silva, M. A., & Paula, P. B. (2002). Estudo dos partos e nascidos vivos de mães adolescentes e adultas jovens no Município de Feira de Santana, Bahia, Brasil, 1998. *Cadernos De Saúde Pública*, 18(3), 715-722. doi: 10.1590/S0102-311X2002000300020
- D'Orsi, E., & Carvalho, M. S. (1998). Perfil de nascimentos no Município do Rio de Janeiro: uma análise espacial. *Cadernos De Saúde Pública*, 14(2), 367-379. doi: 10.1590/S0102-311X1998000200021
- Eden, K., & Kefalas, M. (2005). *Promises I can keep: why poor women put motherhood before marriage*. Berkeley: University of California Press.
- Folle, E., & Geib, L. T. C. (2004). Representações sociais das primíparas adolescentes sobre o cuidado materno ao recém-nascido. *Revista Latino-Americana de Enfermagem*, 12(2), 183-190.
- Fraser, A. M., Brockert, J. E., & Ward, R. H. (1995). Association of young maternal age with adverse reproductive outcomes. *New England Journal of Medicine*, 332, 1113-1117.
- Freitas, D. (1981). *Palmares, a guerra dos escravos*. 3rd ed. Rio de Janeiro, RJ: Graal.
- Gama, S. D., Szwarcwald, C. L., Leal, M. C., & Theme Filha, M. M. (2001). Gravidez na adolescência como fator de risco para baixo peso ao nascer no Município do Rio de Janeiro, 1996 a 1998. *Revista De Saúde Pública*, 35(1), 74-78.

- Geronimus, A.** (2003). Damned if you do: culture, identity, privilege, and teenage childbearing in the United States. *Social Science & Medicine*, 57(5), 881-893.
- Ginsburg, F., & Rapp, R.** (1991). The politics of reproduction. *Annual Review of Anthropology*, 20(1), 311-343.
- Ginsburg, F. D., & Rapp, R.** (1995, 2ed. 2007). *Conceiving the new world order: the global politics of reproduction*. Berkeley, Calif. [u.a.: Univ. of California Press].
- Governo Federal do Brasil**, "Lei 12.255/2010 06/15/2010." Retrieved from legislação.planalto.gov.br/legisla/legislacao.nsf/Viw_Identificacao/lei%2012.255-2010?OpenDocument. 13 Dec 2011.
- Hotz, McElroy, & Sanders** (2005). Teenage childbearing and its life cycle consequences: Exploiting a natural experiment. *Journal of Human Resources*, 40(3): 683-715.
- Jorgensen, S. R.** (1993). Adolescent pregnancy and parenting. In T. P. Gullotta, G. R. Adams & R. Montmayor (Orgs.), *Adolescent sexuality* (pp. 103-140). Newbury Park: Sage.
- Klein, H. S.** (1987). A demografia do trafico Atlantico de escravos para o Brasil. *Estudos Economicos*, 17(2), 129-149.
- Lawlor, D., & Shaw, M.** (2002). Too much too young? Teenage pregnancy is not a public health problem. *International Journal of Epidemiology*, 31(3), 552-553.
- Levandowski, D. C., & Piccinini, C. A.** (2002). A interação pai-bebê entre pais adolescentes e adultos. *Psicologia: Reflexão E Crítica*, 15(2), 413-424. doi: 10.1590/S0102-79722002000200018
- Levandowski, D. C., Piccinini, C. A., & Lopes, R. S.** (2008). Maternidade adolescente. *Estudos De Psicologia (Campinas)*, 25(2), 251-263. doi: 10.1590/S0103-166X2008000200010
- LeVine, R. A.** (1988). Human parental care: Universal goals, cultural strategies, individual behavior. *New Directions for Child and Adolescent Development*, 1988(40), 3-12. doi: 10.1002/cd.23219884003
- Lincoln, R., Jaffe, F. S., & Ambrose, L.** (1976). *11 million teenagers: what can be done about the epidemic of adolescent pregnancies in the United States*. New York: Alan Guttmacher Institute.
- Lubarsky, S. L., Schiff, E., Friedman, S. A., Mercer, B. M., & Sibai, B. M.** (1994). Obstetric characteristics among nulliparous under age 15. *Obstetrics and Gynecology*, 84, 365-368.
- Luster, T., & Mittelstaedt, M.** (1993). Adolescent mothers. In T. Luster & L. Okagaki (Orgs.), *Parenting: an ecological perspective* (pp. 69-99). Hillsdale, New Jersey: Lawrence Erlbaum.
- McAnarney, E. R., Lawrence, R. A., Aten, M. J., & Iker, H. P.** (1984). Adolescent mothers and their infants. *Pediatrics*, 73(3), 358-362.
- Mercês Bahia Bock, A.** (2004). A perspectiva socio-historica de leontiev e a critica a naturalizacao da formacao do ser humano: Adolescencia em questao. *Cadernos CEDES*, 24(62), 26-43.
- Olausson, P. O., Cnattingius, S., & Goldenberg, R. L.** (1997). Determinants of poor pregnancy outcomes among teenagers in Sweden. *Obstetrical & Gynecological Survey*, 89(3), 451-457. doi: 10.1097/00006250-199703000-00027
- Oliveira, M. W.** (1998). Gravidez na adolescência: dimensões do problema. *Caderno CEDES*, 19(45), 48-70.
- Osofsky, J. D., Hann, D. M., & Peebles, C.** (1993). Adolescent parenthood: risks and opportunities for mothers and infants. In *Community Health*.
- Paraguassú, A. B., Costa, M. O., Nascimento Sobrinho, C. L., Patel, B. N., Freitas, J. D., & Araújo, F. D.** (2005). Situação sociodemográfica e de saúde reprodutiva pré e pós-gestacional de adolescentes, Feira de Santana, Bahia, Brasil. *Ciência & Saúde Coletiva*, 10(2). doi: 10.1590/S1413-81232005000200015

- Ribeiro, E. R. O., Barbieri, M. A., Bettiol, H., & Silva, A. A. M.** (2000). Comparação entre duas coortes de mães adolescentes em municípios do Sudeste do Brasil. *Revista de Saúde Pública, 34*(2), 136-142.
- Satin, A. J., Leveno, K. J., Sherman, M. L., Reedy, N. J., Lowe, T. W., & McIntire, D. D.** (1994). Maternal youth and pregnancy outcomes: Middle school versus high school age groups compared with women beyond the teen years. *American Journal of Obstetrics and Gynecology, 171*(1), 184-187.
- Smith, G. C., & Pell, I. P.** (2001). Teenage pregnancy and risk of adverse perinatal outcomes associated with first and second births: Population based retrospective cohort study. *British Medical Journal, 323*, 476-479.
- Yazlle, M. E. H., Mendes, M. C., Patta, M. C., Rocha, J. S. Y., Azevedo, G. D., & Marcolin, A. C.** (2002). A adolescente grávida: alguns indicadores sociais. *Revista Brasileira de Ginecologia e Obstetrícia, 24*(9), 609-614.

SELF-ASSEMBLED MONOLAYERS:
A Study of Si and GaAs Surface Modification

simbarashe marufu

Simbarashe Marufu is an undergraduate physics and mathematics major. He is graduating in spring 2013 as a member of the Ronald E. McNair Postbaccalaureate Achievement Program. He plans to attend graduate school while pursuing a PhD in either Quantum Physics or Solid State Physics. After graduate school, he hopes to have a career studying quantum phenomenon on nano scale surfaces either in academia or industry. His research was conducted under the guidance of Dr. Theodosia Gougousi and Lewang Ye, a graduate student working with Dr. Gougousi. Many thanks are due to both. The McNair Program was also a tremendous source of support and deserves thanks. Marufu would also like to thank the NIH for providing funds in the form of a research grant.

In today's technology driven world, the demand for smaller devices with considerable amounts of processing power is greater than ever before. As these devices become smaller, circuit elements on a nano scale must be created in order to provide the necessary processing power in a very confined volume. Of particular interest are the dielectrics used in capacitors. For the past several decades, SiO_2 (silicon dioxide) has become the substance of choice for these dielectrics; however, as circuit elements have become smaller, SiO_2 has become less suitable to use. Surface imperfections become a major problem on the nano scale and have rendered the use of SiO_2 gate oxides impractical. One method of producing high quality materials is through a process called atomic layer deposition (ALD). This process requires surface modification using self-assembled monolayers (SAMs). A self-assembled monolayer is a layer of molecules arranged in such a way that the head groups are attached to the substrate and the functional groups are aligned on the surface. This surface modification was the focus of my research. The most captivating aspect of the research I conducted was how closely SAMs mimic natural processes. Equally as interesting is the potential SAMs have for widespread technological applications.

INTRODUCTION

OVER THE PAST decade, considerable attention has been given to the growth and development of nano film structures. This increased attention is largely due to the growing demand for materials with more ubiquitous surface structures. Materials with this property are essential in today's technological devices that demand smaller and smaller components. One way of producing such surfaces is through the application of self-assembled monolayers. A self-assembled monolayer (SAM) is an organized coating of molecules on a surface such that each molecule is bonded to the underlying substrate at one end and unbonded at the other end. The group bonded to the surface is typically referred to as the head group and the unbonded group is referred to as the functional group.

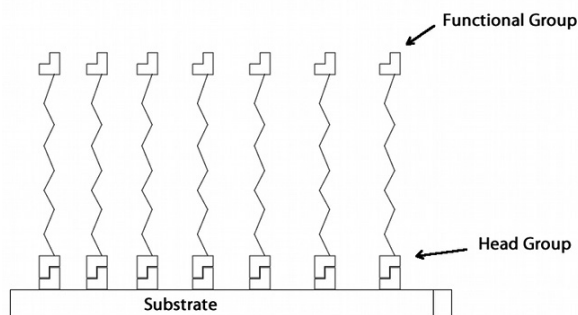


FIG. 1 A general representation of a layer of SAMs (Figure redrawn by author. Original figure from reference 3).

The ability of these SAMs to survive for prolonged periods of time in ambient exposure and under varying temperature conditions is of tremendous importance. In order for SAMs to be applied in processes such as atomic layer deposition or micro contact printing, they must have longevity.

The substrates examined were gallium arsenide (GaAs) and silicon (Si), and the molecule used for SAM formation was octadecyltrichloro-silane (OTS). The focus of the study was on the assembly of OTS molecules on Si surfaces and their viability under ambient exposure at room and elevated temperatures.

One way of examining the degeneration of SAMs is to evaluate whether or not the surface hydrophobicity changes. A surface is said to be hydrophobic if it does not absorb water and any water placed

on such a surface beads. Conversely, a hydrophilic surface readily absorbs water. A hydrophilic surface will undergo a reaction with OTS resulting in a hydrophobic surface. We exposed this hydrophobic surface to varying temperatures and subsequently tested to see if it had degraded, or if the surface was viable and had maintained hydrophobicity. We also tested to see if SAM formations remain bonded to the surface after prolonged ambient air exposure. We hypothesized that hydrophobic SAMs on a Si or GaAs substrate will not degenerate when exposed to low temperatures but will gradually begin to degenerate as the temperature is increased. We also hypothesized that the SAMs will begin to degenerate over extended periods of ambient exposure leaving a hydrophilic surface. These surfaces were primarily monitored using Fourier transform infrared spectroscopy (FTIR).

BACKGROUND

Surface Formation

SAMs ARE TYPICALLY placed in four distinct groups, each with different chemical properties and each reactive to different substrate surface groups. The first class of SAMs is organo-sulfur compounds; these compounds typically bond on late transition metals. Fatty acids tend to bond on metal oxides, while alkylphosphonic acids bond to a large variety of metal and nonmetal surfaces. The last classification belongs to organosilicon compounds, which bond to metal and nonmetal oxides. [11] OTS is an organosilicon compound that we bonded to the metal oxide SiO_2 . Octadecyltrichloro-silane(OTS) is an organometallic chemical with the formula: $\text{CH}_3(\text{CH}_2)_{17}\text{SiCl}_3$.

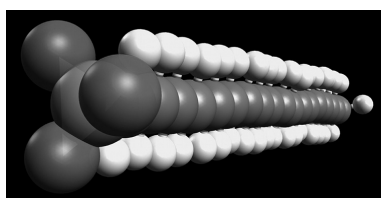
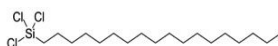


FIG. 2 General representations of an OTS molecule (Figure redrawn by author. Original figure from reference 7)

The polar head group for OTS is SiCl_3 and the tail of the body of OTS is a long chain-alkyl group. The head group reacts with the substrate surface structure resulting in the formation of the SAMs. This process occurs spontaneously and results in the release of energy which leaves behind stable film formation. This surface aggregation largely results from covalent bond formation and Van de Waals interactions between the OTS hydrocarbon chains. [11]

SAM Defects

THE TOPOLOGY OF a substrate covered with SAMs is largely uniform, but in small density, exhibits defects. If the density of the defects is large enough, then the defects aid in degenerative processes. Furthermore, a large density of defects also makes SAMs inapplicable in processes such as atomic layer deposition (ALD). Although it is impossible to eliminate all defects, it is critical to minimize their density. Logically, it then follows that an evaluation of the main causes of these defects is necessary to determine ways of minimizing them. [5]

Causes of defects are characterized as either external, intrinsic, or a mixture of both. External causes are environmental, whereas intrinsic causes are largely structural. Intrinsic defects tend to be caused by the thermodynamics of formation and the topology of the substrate. A substrate with many abrupt topological changes promotes defects, therefore to minimize defects it is prudent to use a substrate with a smooth topology. [5] External defects are influenced by cleanliness of the substrate, methods of preparation, and the purity of the adsorbents. A cleaner substrate limits anomalous surface reactions with contaminants; this results in greater availability of bonding sites and allows for uniform SAM development. [11] The cleaning procedures used are detailed in the methodology section. The method of immersion is reliable for producing a dense SAM and was used to saturate the substrate. Pristine vials, syringes, and careful handling were used to guarantee purity of adsorbents.

Fourier Transform Infrared Spectroscopy (FTIR)

LIKE ANY INFRARED spectrometer, FTIR works by exposing a sample to many wavelengths of electromagnetic radiation. The sample then absorbs radiation with a particular wavenumber that makes it possible to determine the sample's molecular composition. Once the FTIR spectrometer collects relevant information a spectrum is then generated.

FTIR can be particularly useful because it functions over a wide range of wave numbers, which allows for the collection of more data on the sample composition. [12] The FTIR spectrometer also allows for the collection of information on vibrational and rotational modes of many compounds. With respect to a solid sample, each peak corresponds to vibrational and rotational states within the sample. Wider peaks indicate substances that are on the surface but are not bonded. FTIR is also used to determine the photoconductivity of a sample and to also determine its Raman scattering.

This information can also be used to determine surface composition and many other chemical properties including electrical conductivity and hydrophobicity. [7]

Applications

SAMs ARE ABSOLUTELY essential in ALD. Atomic layer deposition is a subfield of chemical vapor deposition (CVD), which is a state-of-the-art technique for producing high quality thin films through repeated exposure. What makes ALD particularly special is the fact that it is a cyclical process that results in a surface termination that is identical to the original surface termination.

SAMs are also vital in molecular layer deposition (MLD). MLD is very similar to ALD; it is a process that uses long organic molecular chains instead of short inorganic molecules for surface functionalization. Thin films deposited using this process can have great technological applications due to the properties of organic molecules. Circuit elements constructed with organic compounds have the benefit of flexibility and could possibly pave the way for next generation consumer electronics. [5]

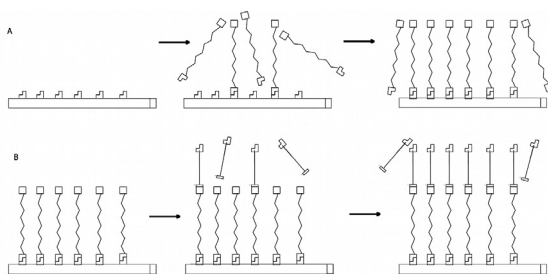


FIG. 3 A general representation of molecular layer deposition on a surface (Figure redrawn by author. Original figure from reference 5). The application of precursors across the surface of a substrate results in surface functionalization. Continued use results in material growth that is dependent upon temperature, pressure, time, and density of precursor.

In the above image, the surface is exposed to a precursor and the subsequent reaction results in a new surface termination. The excess precursor molecules are then purged and exposed to a secondary precursor resulting in the original surface functional group. This process is repeated multiple times to produce a film. The rate of film growth is highly dependent upon temperature. [11]

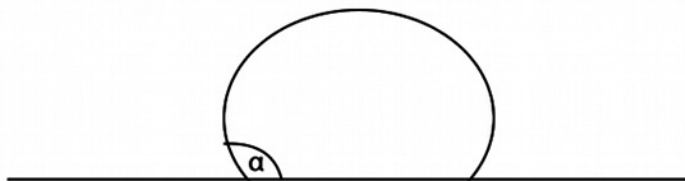


FIG. 4 A water drop on a surface with a contact angle α (Figure redrawn by author. Original figure from reference 7)

One test for the hydrophobicity of a surface is to use contact angles (represented by the angle in the above image). If a water droplet is placed on a surface, it makes an angle with the surface. The resulting angle can then be used to judge the hydrophobicity of the surface. The greater the contact angle the more hydrophobic the surface. A surface with a contact angle greater than 90° is regarded as hydrophobic. [12] This is because a contact angle greater than 90° shows a repulsive force between the surface and the liquid. Alternatively a surface with a contact angle less than 90° shows an attractive surface-liquid interaction and therefore such surfaces are regarded as hydrophilic. [8]

For a wide range of applications it is often useful to be able to block ALD chemistry on parts of a surface. This gives greater control to the experimentalist and has been found to be useful for the development of an area selective ALD process. [1] One way of achieving this area-selective ALD is through selective surface modification. Prior to deposition, a portion of the surface is covered with suitable SAMs, resulting in an inert surface area that will not react with the ALD precursor. [3] Using ALD, only the uncovered portion of the surface experiences growth and HfO_2 film grows selectively on that area. The area covered with SAMs is unreactive and a film does not grow there.

METHODOLOGY

OTS Preparation

THE OTS WAS purchased from Acros Organics. Given that OTS is amphiprotic with a polar head group, it reacts violently to water and air. This reactivity led to the use of anhydrous hexane to dilute the solution and also led to the use of a glove box to create a moisture free environment. It is worth noting that OTS is not reactive with hexane or nitrogen and this was crucial to our experiment. A 15 mM solution was prepared with 10 ml of hexane and 0.06 ml of OTS. Once OTS solutions were prepared, they were handled and sealed in clean vials with aluminum caps and a Teflon septum using a 20 mm head-size crimper.

Substrate Preparation

400MM SILICON AND gallium arsenide wafers were cleaved into 2-3 cm² pieces. These wafers were then placed in JT Baker proprietary cleaning solution for five minutes. This was done to remove large scale contaminants and for degreasing. The wafers were then subsequently placed in deionized water for five minutes. Nitrogen gas was then used to dry and minimize moisture on the silicon and gallium arsenide wafers.

SAM Assembly Using Immersion

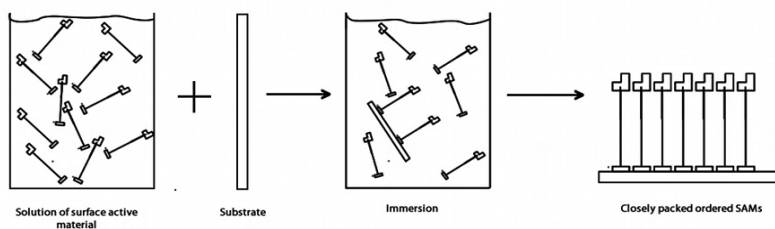


FIG. 5 A general representation of SAM assembly on a surface (Figure redrawn by author. Original figure from reference 11)

The prepared substrates were placed in the prepared solutions while still in the nitrogen environment. The solution was then subsequently crimped and removed from the nitrogen environment. Covalent bonding and intermolecular Van de Waals interactions resulted in SAM formation. Rate of formation was dependent upon temperature, purity of solution, cleanliness of substrate, and concen-

tration of solution. After at least an hour of formation, the substrate was decrimped and removed from the solution and immediately dried using nitrogen gas. In order to see the effects of hexane on FTIR imaging the samples were sometimes cleaned in methanol and water, and subsequently dried in nitrogen gas.

Testing Methodology for Ambient Exposure and Varying Temperatures

FOR AMBIENT EXPOSURE, samples with SAM formations first had reference spectra collected using FTIR spectroscopy, then were placed in an environment at STP for varying lengths of time. Spectra were collected again after the exposure. The spectra prior to the exposure were then compared to the spectra after the exposure.

In order to test for SAM degradation due to high temperature exposure, reference spectra were first collected on samples with SAM formations; subsequently they were placed in an oven at different temperatures for varying lengths of time. More spectra were then obtained, and compared to the reference spectra obtained prior to the high temperature exposure.

In order to examine how different temperatures affected SAM growth rates, clean silicon samples were immersed in OTS solution. Afterward, these samples were placed in environments with different temperatures for one hour before being collected and compared.

The solution used was prepared with hexane and the instruments were cleaned with methane. It is prudent to ask if either hexane or methane would mimic the appearance of OTS on a surface. These chemicals have C-H bonds; therefore it is a valid concern.

Hexane was first to be analyzed and doing so required the use of FTIR microscopy.

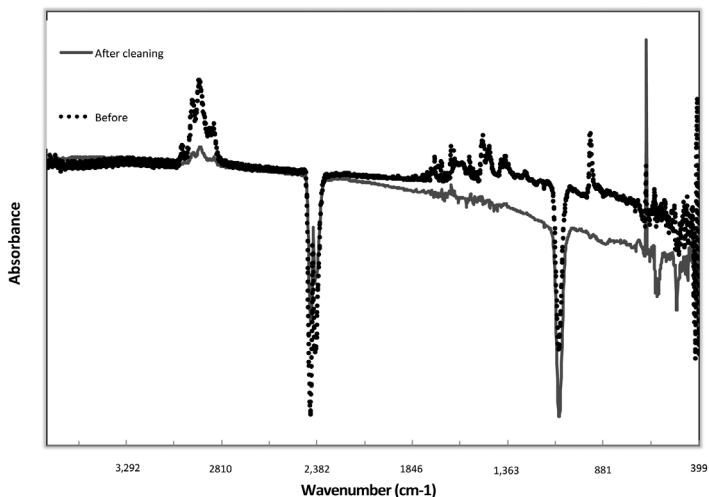


FIG. 6 Relevant peaks are centered on the wavenumber 3000; these peaks correspond to C-H bonds in hexane. They also show the different modes for those bonds.

As shown in Figure 6, when the sample was cleaned and rinsed in deionized water, the overwhelming majority of the hexane was removed clearly; this shows that no bonding took place between the silicon oxide and the hexane. Given that hexane was not bonded, the peaks produced were shallow relative to those produced by OTS (see below). It is also reasonable to believe that OTS bonds to active sites and subsequently blocks and repels hexane given that the end groups have identical polarizations. The peaks produced clearly do not mimic those produced by bonded OTS.

Methane was next to be evaluated; the results were similar. The only difference between hexane and methane was that the peaks for methane were much shallower than those of hexane, and often were barely noticeable. After rinsing and drying, the peaks vanished. Clearly the use of methane was not a valid source of concern.

RESULTS

Air Ambient Exposure

SILICON SAMPLES WITH OTS SAM formation were placed in ambient exposure for four different lengths of time. Samples were exposed for one day, two days, three days, and five weeks before spectra were collected.

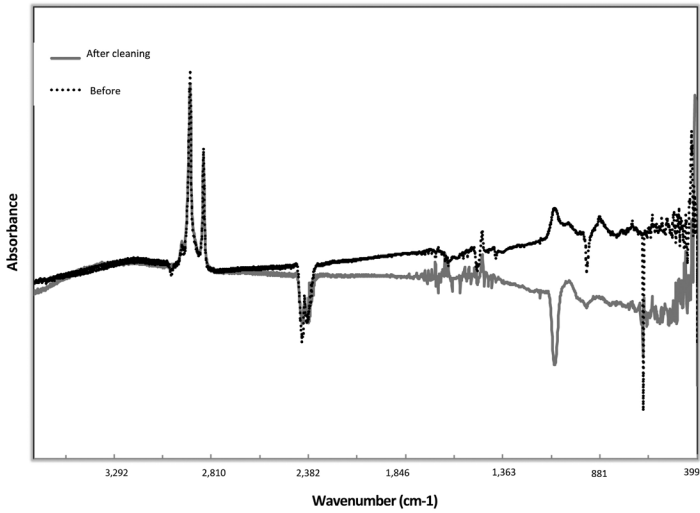


FIG. 7 These spectra were collected before and after one day of ambient exposure. Relevant peaks are centered around the wavenumber 3000, these peaks correspond to C-H bonds They also show the different modes for those bonds

The spectra are practically identical. The peaks are equally sharp and have the same area under the curve, a clear indication that the amount of OTS bonded to the surface did not change.

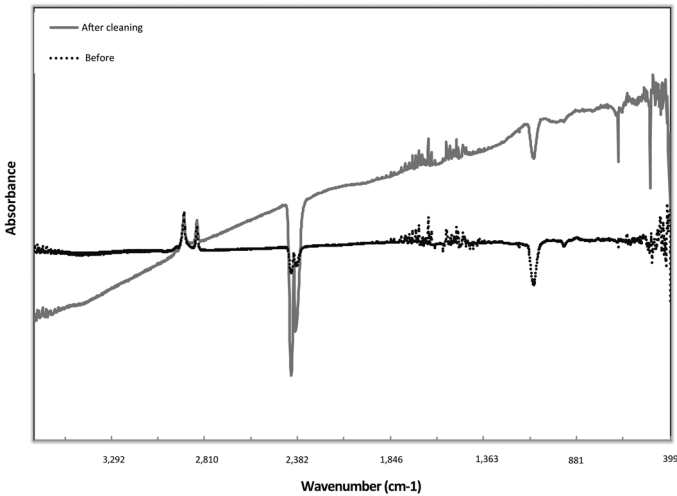


FIG. 8 These spectra correspond to five weeks of ambient exposure. Relevant peaks are centered around the wavenumber 3000, these peaks correspond to C-H bonds They also show the different modes for those bonds. These bonds are indicative of the presence of OTS on the surface. Sharp peaks show that OTS is bonded to the surface.

The spectra are practically identical. The peaks are equally sharp and have the same area under the curve. The OTS on the surface clearly did not degrade even after such an extended period of time. GaAs samples with OTS SAM formation were also placed in ambient exposure for two weeks with similar results.

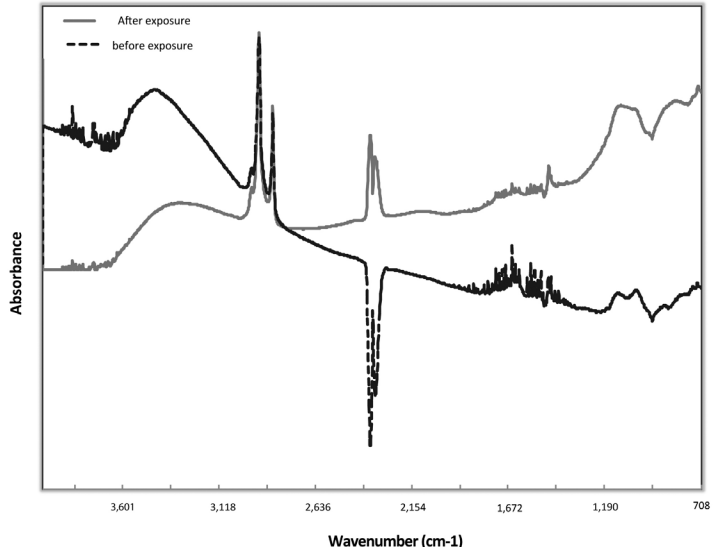


FIG. 9 The shape and size of the peaks shows very little change. This suggests that ambient exposure did not degenerate OTS. The following spectra correspond to before and after two weeks of ambient exposure.

The spectra are almost identical. The peaks are equally sharp and have the same area under the curve. Therefore, the OTS on the surface did not degrade.

We can conclude that ambient exposure has no effect on the integrity of SAM formations on both gallium Arsenide and silicon at room temperature. Even after five weeks there was no degradation.

Temperature Exposure

SILICON SAMPLES WITH OTS formations were placed in a furnace at 200°C and 250°C temperatures for varying lengths of time. Samples were exposed for one hour, two hours and three hours before spectra was collected. For the 250°C exposure, the following spectra were collected after an interval of two hours.

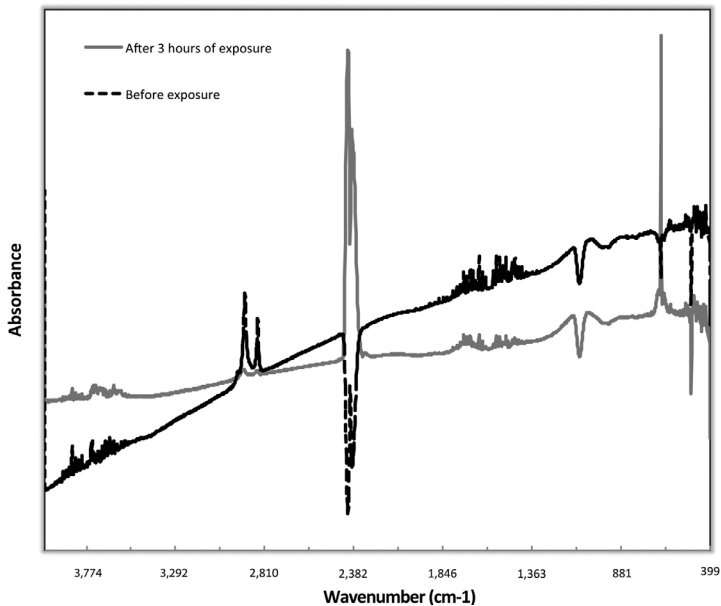


FIG. 10 SAM degradation is obvious on the surface; only trace amounts of OTS are left. For the 250°C exposure the following spectra were collected after an interval of two hours.

Only residual amounts of OTS were left on the surface after the exposure. The OTS completely degenerated and similar results were found for the one hour and three hour exposures.

Temperature Dependence of Formation

SILICON SAMPLES WERE placed in OTS solution at 60°C to determine if rate of SAM growth changed due to the temperature change. The results were dramatic.

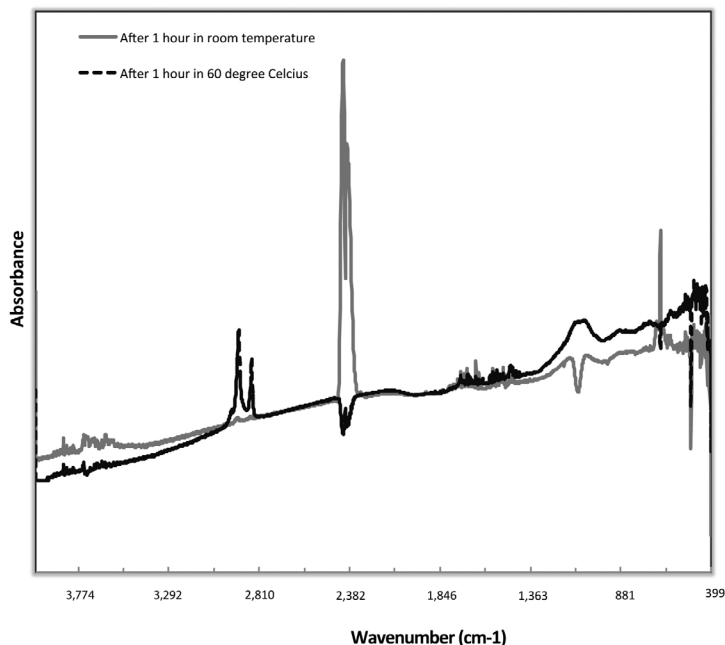


FIG. 11 Relevant peaks are centered around the wavenumber 3000. The red line corresponds to after one hour of 60° exposure, and the blue line corresponds to a spectrum collected after one hour of room temperature exposure.

After only an hour in a 60°C solution, SAMs were fully formed, compared with room temperature where SAM formation was hardly noticeable. It was expected that higher temperatures would promote faster SAM development, but certainly not to such a great extent. Clearly SAM development is much faster at higher temperatures.

CONCLUSIONS

THE RESULTS FOR ambient exposure were definitive. The spectra for one day, two days, three days, one week, and five weeks of exposure all showed no signs of SAM degradation. These results were independent of the substrate used. Neither gallium arsenide nor silicon OTS SAM formations degraded.

The results for temperature exposure were also very illuminating. The SAM formations were initially exposed to temperatures of 50°C, 60°C, and 90°C with no sign of degradation after three hours. However, when the samples were exposed to 250°C for one hour, two hours, and three hours, SAM layers completely disintegrated. At

200°C, after one hour of exposure, there was little evidence of degradation. Therefore, it can be concluded that OTS SAM formations on Si substrates begin to show significant signs of degradation between 200°C and 250°C when exposed for an hour. It can also be surmised that temperatures higher than 250°C also achieve complete degradation of SAM formations.

Pertaining to the temperature dependence of SAM formation, the results were also quite expressive. After Si wafers were placed in a 60°C environment while in solution for one hour, SAM assembly was visibly more profound than on wafers in room temperature. We can conclude that higher temperatures promote greater SAM assembly rates. The results of our experimentation were largely reflective of our hypothesis; thus it can be said that our experiment was success

REFERENCES

- 1 **Bent, S.F, Chen, R, Kim, H, & McIntyre, P.C.** (2004). Self-assembled monolayer resist for atomic layer deposition of hfo₂ and zro₂ high-k gate dielectrics. *APPLIED PHYSICS LETTERS*, 84(20), 4017-4019.
- 2 **Bent, S.F, Chen, R, Kim, H, McIntyre, P.C, & Porter, D.W.** (2005). Achieving area-selective atomic layer deposition on patterned substrates by selective surface modification. *APPLIED PHYSICS LETTERS*, 86(1919), 1-3.
- 3 **Bent, S, & Chen, R.** (2006). Chemistry for positive pattern transfer using area-selective atomic layer deposition**. *ADVANCED MATERIALS*, 10(1002), 1086-1090.
- 4 **Bertone, J.F, Colvin, V.L, Hwang, K.S, & Jiang, P.** (1999). Single-crystal colloidal multilayers of controlled thickness. *Chem. Mater.*, 11, 2132-2140.
- 5 **Dameron, A.A, George, S.M, & Yoon, B.** (2008). Surface chemistry for molecular layer deposition of organic and hybrid organic-inorganic polymers. *ACCOUNTS OF chemical research*, 1-11.
- 6 **El-Komy, G.M, Sakr G.B., Salem, A.M, & Yahia, I.S.** (2011). Optical properties of thermally evaporated bi₂se₃ thin films treated with agno₃ solution. *Surface and Coatings Technology*, 205(12), 3553-3558.
- 7 **Lacis, J.W.** (2009). Atomic layer deposition of ti₀₂ on si and gaas substrates using tdm₂tiand h₂o precursors, (Master's Thesis) University of Maryland Baltimore County, Baltimore, MD.
- 8 **Link, D.O.** (2007). Selective deposition of metal-oxides on silicon using self-assembled monolayer resists, (Master's Thesis) University of Maryland Baltimore County, Baltimore, MD.
- 9 **Freire, Jr F.L, Maia da Costa, M.E.H, Rebollo Plata, B., & Sanchez, C.M.T.** (2011). Titanium diboride thin films produced by dc-magnetron sputtering: structural and mechanical properties. *Surface and Coatings Technology*, 205(12), 3698-3702.
- 10 **Hyde, K, Peng, Q, Spagnola, J, Spontak, R & Sun, X.** (2007). Atomic layer deposition on electrospun polymer fibers as a direct route to al₂o₃ microtubes with precise wall thickness control. *NANO Letters*, 3 251-254.
- 11 **Kim, J.H, Moholkar, A.V, Rajpure, K.Y., & Shinde, S.S.** (2011). Structural, morphological, luminescent and electronic properties of sprayed aluminium incorporated iron oxide thin films. *Surface and Coatings Technology*, 12(12), 3567-3577.
- 12 **Lee, J.H, McCormick, A, Schwartz, B.T, & Sechrist, Z.A.** (2006). Modification of opal photonic crystals using al₂o₃ atomic layer deposition. *Chem. Mater.*, 18, 3562-3570.

NATO

Identity, Role, and Function in the Post-Cold War Era

joshua hunter

Joshua Hunter graduated from UMBC in 2011 with a B.A. in Political Science, and minors in international affairs and modern languages and linguistics (French). Following graduation, he began a grants and research internship at Mines Advisory Group (MAG) America. He would like to pursue a career in the nonprofit sector, particularly within the area of conflict recovery. He thanks his mentor, Dr. Brigid Starkey, for her support and encouragement throughout the duration of the project; Dr. Devin Hagerty for additional guidance; Dr. Cynthia Hody; the faculty, staff, and students of the Department of Political Science; the editors of the *UMBC Review*; and his friends and family.

I chose to write about the North Atlantic Treaty Organization (NATO) in order to explore American foreign policy and its influence on NATO in the conduct of international relations. I began the research process with a number of preconceptions about the subject, but I challenged these as I learned more about the topic. After much research, I refined my topic to argue that NATO is a viable actor within the international community for two reasons: (1) its member states view it as an essential component to their individual security, and (2) it has successfully involved itself in a number of nontraditional crises in the post-Soviet security environment. This paper contends that NATO maintains a strong role in the contemporary international system. Although many have argued that NATO is an irrelevant alliance in the post-Cold War era, the perspective offered here is a useful contribution to the discourse of NATO as a security organization.

INTRODUCTION

THE NORTH ATLANTIC Treaty Organization (NATO) was established in response to the emerging Soviet threat in the wake of World War II. The alliance created a successful balance to this perceived threat by effectively deterring further Soviet expansion or aggression in Europe. This job of deterrence was not disputed as NATO's primary function during the period from 1949 to 1991. However, once the Eastern bloc surprisingly disappeared — and did so peacefully — there were many predictions that NATO would dissolve, which had been the pattern with previous alliances and shifting international configurations. However, this has not been the case. Twenty-one years after the demise of NATO's primary adversary, the alliance continues to operate. This raises the question of why NATO continues to exist.

The literature on traditional alliance behavior does not offer an explanation of NATO's perseverance beyond the Cold War against all odds.¹ This paper argues that NATO is relevant to the post-Soviet security environment for two major reasons: NATO now successfully serves a number of important functions for its member states, and NATO continues to maintain a viable alliance-based identity despite frequently heard rumblings that it merely provides cover for American foreign policy endeavors. This paper opens with a discussion of theoretical perspectives through which NATO will be analyzed. It then addresses NATO's roles and functions in the international system. Finally, it examines NATO's identity in terms of community theory, burden-sharing, and public support. These discussions identify concrete means by which NATO has asserted its relevance in the contemporary international environment.

NATO AND THEORETICAL PERSPECTIVES

TRADITIONALLY, ALLIANCE BEHAVIOR has been explained using the realist paradigm, the dominant framework in analysis of international relations. However, constructivism, which includes the roles of identity, context, and institutions as explanatory factors for actor behavior, often provides a more complete analysis of state actions than does realism. Nonetheless, NATO was formed at a time when international relations were best explained by realism. For this reason, it is important that the utility of this theoretical perspective be analyzed with regard to NATO's role during the contemporary period.

Realism posits that states are primarily concerned with survival and relative power.² Therefore, a state would be expected to enter into an alliance with another state or multiple states if it cannot achieve its goals alone.³ This inter-dependency results in a redistribution of power within the international system. Stephen Walt explains that states create alliances in one of two ways: “when entering an alliance, states may either balance (ally in opposition to the principal source of danger) or bandwagon (ally with the state that poses the major threat).”⁴ States may choose to bandwagon when they fear the power of a strong state or alliance or when there is no alternative.⁵ States may choose to balance in order to prevent the dominance of one state over all the other parts within the system while remaining autonomous.⁶

It is useful to distinguish the role of *power* and *threat* in regards to forming alliances: “[r]ather than allying in response to power alone, it is more accurate to say that states will ally with or against the most *threatening* power.”⁷ Such a distinction is important because raw capabilities can be threatening if other states view such capabilities as supporting what Walt calls “offensive intentions.”⁸ Given these features of realism, its adherents expect direct conflict between opposing powers. However, this direct conflict did not occur during the Cold War. One is thus confronted with a significant challenge to the realist framework. A further challenge to realism is presented by NATO’s behavior *following* the Cold War. At the point of communist collapse, realists expected NATO to dissolve because it ceased to have a defining existential purpose.⁹ Contrary to predictions, NATO did not identify a new rising power to take the role of its threat (Japan, Germany, or China, for example).¹⁰ Realism explains NATO’s origins, but it cannot sufficiently explain NATO’s persistence. It is possible that NATO is an exception to typical alliance behavior.¹¹

Where realism fails to account for NATO’s persistence, constructivism provides a more useful framework. This perspective holds that there are multiple types of actors of significance within the international system. In addition, the “interests and identities of actors in world politics are malleable,” and the social context in which actors operate is the product of their “actions and interactions.”¹² The most important contribution that constructivism lends to the current analysis is the idea that interests and identities can change. Although NATO allies no longer needed to respond to the Soviet threat, the alliance proved to be useful in response to other

threats. NATO has been able to posture itself as a responder to non-existential threats (threats which are not expected to result in the destruction of a member state or the alliance itself). Identity change has also been evident in NATO's ability to work alongside other types of non-state actors toward the goal of enhancing European security together. Constructivism allows for the role of identity in the actions of states, but it does not specifically account for cause and effect relationships. It does suggest, though, that identity plays a role in the social surroundings in which states exist. This is not a one-way relationship — states themselves can have an influence on the context of the environment in which they operate.¹³ This relationship between environment and actor explains not only NATO's origins in reaction to perceptions of a growing Soviet threat in the 1950s but also its current involvement in combating nontraditional threats.

Constructivism further allows for recognition of the importance of institutions. Broadly put, "institutions are settled or routinized practices established and regulated by norms."¹⁴ This notion is important because it indicates that states will adhere to such norms rather than act solely in their own self-interest (if the two are not wholly incompatible). In addition, institutions can help states overcome challenges of conducting international relations by easing information-sharing and increasing transparency.¹⁵ Celeste A. Wallander argues that there is a rational calculation that plays a role in state decisions to maintain norms and institutions. Simply put, it is easier for states to operate within established procedures and power relations than to create new ones.¹⁶ In fact, attempts to create alternatives to NATO have been unsuccessful for reasons Wallander describes, which will be discussed later.

The two theoretical arguments above appear to privilege states as primary actors in the international system. In terms of realism, states are privileged and carry the most weight. Yet, contrary to this assumption, NATO has demonstrated that an intergovernmental organization can bear significant weight in this contemporary international system. According to constructivism, states will create and maintain institutions to respond to new threats that are considered detrimental to their constituents. Alliances help mitigate problems associated with a self-reliant, anarchical environment.¹⁷ Contrary to responding to solely traditional state threats, NATO has operated in humanitarian interventions and worked to counter transnational terrorist organizations. It has also worked on cyber-terrorism and environmental degradation, both of which have become major

security concerns for actors throughout the international system in the twenty-first century. NATO's specific functions in the post-Cold War era are enlargement, peacekeeping, and nuclear arms control. These functions are discussed in greater detail in the section that follows.

ROLE AND FUNCTION

AS STATED BEFORE, NATO's primary role is to provide security for its member states. During the Cold War, it achieved this through the size and strength of its conventional and nuclear capabilities as deterrents to its Soviet counterparts.¹⁸ Today, the Soviet-era threats have diminished. The post-Cold War period has required NATO to reassess the international security environment and adapt to the changes that have since occurred. This reassessment has entailed an active approach to security, requiring a posture of involvement rather than solely one of deterrence. This shift introduces an underlying theme: the distinction between types of threat and sources of threat. Types of threat are either means (tactics) of attacking another actor or the destructive capabilities of an actor. For example, specific types of threat are transnational terrorism, cyberterrorism, regional instability, and weapons of mass destruction proliferation. On the other hand, sources of threat are regarded as specific actors which employ these various types. Sources of threat can be actors in the international system, namely states and transnational groups of a wide variety. The fact that NATO considers these actors as legitimate threats represents a significant shift from Cold War thinking. This shift may have occurred because NATO faces no ideological or hostile rivals equivalent to the rivalry during the Cold War, nor does it face an existential threat remotely similar to that during the Cold War. A diminished likelihood of great power conflict at the end of the Cold War allowed NATO to concern itself with non-existential threats and other types of conflict that developed, which explains the increasing importance of non-state threats in the post-Soviet security environment.¹⁹

NATO has taken an active approach to the contemporary security environment. Given the dynamic nature of modern threats, NATO has been forced to pursue means other than deterrence to address them. Deterrence cannot sufficiently deal with each type of threat or even each source of threat. For instance, conventional means of deter-

rence will not prevent or discourage cyber-terrorism or transnational terrorism. Cases of each have occurred against NATO member-states regardless of its traditional deterrent capability (for example, al-Qaeda's explicit targeting of NATO members for terrorist attacks, including Spain and the United Kingdom).

NATO stated in its most recent Strategic Concept that it "does not consider any country to be its adversary," suggesting the absence of traditional security concerns.²⁰ However, as its involvement in Afghanistan demonstrated, a weak or failed state can be a source of alliance insecurity. Afghanistan, as a failed state, "contributes to regional insecurity, weapons proliferation, narcotics trafficking, and terrorism."²¹ Due to the weak condition of the Afghan government prior to September 11, 2001, al-Qaeda was able to thrive and plan the 9/11 attacks against the United States.²² The opium trade, which has helped fund the Taliban since the American invasion and NATO's subsequent involvement, provides yet another challenge that NATO must face.²³

This section explores specific functions that NATO uses to address nontraditional types of threats that are pervasive in the current international security environment. These functions have served the "three essential core tasks" which comprise NATO's role: to maintain "collective defense," to provide "crisis management," and to facilitate "cooperative security."²⁴ The notion of "crisis management," by its name alone, indicates a major departure from Cold War posturing.

NATO's Scope in Post-Cold War Europe

THE INCLUSION OF former Warsaw Pact states (e.g., Poland, Hungary, Czech Republic) has been an important part of NATO's strategy in the post-Cold War period. NATO has been able to include these countries through its policy of general enlargement, which the Partnerships for Peace Program (PfP) has facilitated. Partnerships may entail full membership to the alliance, although this is not always the case.²⁵ NATO did expand during the Cold War, but such enlargements did not include former adversaries as defined by the Cold War.²⁶

Expansion has increased NATO's influence closer to Russia's borders, though with much caution due to inherent, lingering tension since the end of the Cold War. However, enlargement through

partnerships and new memberships has allowed the alliance to greatly improve its relationship with former Soviet states.

In 1995, NATO conducted the “Study on NATO Enlargement” to examine the potential for additional post-Cold War growth. This study indicated the types of security contributions that future members were expected to make. Perhaps most significantly, NATO called for “cooperation” and “increasing transparency.”²⁷ Enlargement in theory, along with joint military exercises, allow greater freedom of movement for NATO troops.²⁸ NATO only invites states to join if those states can offer a means of enhancing NATO security beyond current capabilities.²⁹

Much of NATO’s calculations on enlargement have concerned reducing uncertainty in international affairs, particularly in the European context. One primary concern is to overcome long-standing tension between East and West that extends beyond tensions with Russia.³⁰ NATO’s emphasis on cooperation permits increased predictability of the behavior of its member states and partners as well as increased confidence among the actors. Enlargement, by definition, also reduces the number of nonmembers and therefore potential adversaries on the European continent. This reduction decreases the likelihood of conflict in Europe, which is an obvious, long-standing goal for Western states given Europe’s violent history. In addition, the alliance emphasizes peaceful conflict resolution for its member states. Cooperative actions are not limited to NATO members; partnerships have allowed NATO to engage in “practical bilateral cooperation” with nonmembers.³¹

A key element of NATO’s Partnerships for Peace program and its enlargement process has been an emphasis on democratic reform for new members in the post-Cold War era.³² This reform indicates a major shift from NATO policy regarding membership during the Cold War, since Portugal, one of its founding members, was then a “ruthless dictatorship.”³³ Turkey’s admittance to NATO in 1955 presents another case in which this prerequisite was not enforced.³⁴ NATO states that Cold War enlargement occurred “when strategic considerations were at the forefront of decision-making.”³⁵ This suggests that the higher cause of deterring and defending against the Soviet threat was of greatest concern. Because of the absence of a current existential threat to NATO members, today a loose semblance of democracy must be established as a prerequisite for new members to join the alliance. Again, however,

NATO's primary purpose continues to be the provision of security for its member states.

Though many of the most recent members admitted to the alliance since the end of the Cold War are democracies, Albania is an exception; it was considered only "partly free" by Freedom House at the time of its incorporation into the alliance and after.³⁶ Since Albania was one of the two most recently admitted states to the alliance, this suggests that the most important consideration for membership continues to be security. One scholar suggests the possibility that, "establishing cooperative frameworks and encouraging democratic rule may simply be considered the most cost-effective security strategy" for the alliance.³⁷

Although partnerships and enlargement allow greater troop mobility within NATO's territory, redeployment of troops within newly-admitted states has not occurred. In fact, Anne Applebaum noted in a *Washington Post* column that:

*...the placement of NATO forces and institutions has hardly changed in two decades. The alliance now has 28 members, including almost all of the states that used to be the Warsaw Pact, but the three joint force commands are all still in the south and west of the continent, in Portugal, southern Italy and the Netherlands. American forces are dispersed in odd ways as well. More than 50,000 U.S. troops are based in Germany – a country now surrounded on all sides by NATO allies – while Poland and Norway, countries with long, non-NATO borders, have 100 and 80 respectively.*³⁸

The reasons for the lack of troop redistribution within NATO's territory may be influenced by the implications such redistribution could have, particularly on relations with Russia. Continued NATO expansion toward Russia's borders, including troop redistribution in former Warsaw and Soviet states, could be perceived as an American "neocontainment policy"; this is a common Russian perception of such expansion.³⁹ The redistribution of troops into these areas could heighten tensions. Diplomatically speaking, this tension is not lessened when states formerly in the Russian sphere continue to reorient themselves toward the West in terms of sociopolitical association.⁴⁰ There have already been problems with attempts at reconciling relations between NATO and Russia, and there has been little headway

in terms of solving the root causes of such tension. This result stems from negative discourse and bold actions undertaken by both sides.⁴¹ Russia strongly opposed NATO's intervention in Kosovo in 1999, and the United States and NATO strongly opposed Russia's invasion of Georgia in 2008.⁴² The establishment of the "NATO-Russia Founding Act on Mutual Relations, Cooperation and Security" in 1997 was a major step toward smoothing relations between the two. This led to the creation of the NATO-Russia Council (NRC) five years later, which was meant to facilitate political relations between the two entities.⁴³ However, Russia's involvement in Georgia in 2008 resulted in a cessation of NRC activities.⁴⁴ This severing of ties between NATO and Russia was only recently mended as part of an attempt by the United States to restart relations in 2009 under the Obama Administration.⁴⁵

Russia's concerns alone do not drive the alliance's enlargement and partnerships or even its troop distribution. Yet positive relations with Russia have broader security implications than any potential security gains that would accrue from redistributing troops throughout the territories of new member states. The impact of enlargement on NATO-Russian relations demonstrates that external pressures have complicated expansion beyond internal identity questions. Although the alliance has been successful in meeting goals such as increasing cooperation and transparency, this has not led to a redistribution of conventional forces, particularly within newly admitted states on its newly defined border. The fact that Russia has had major concerns regarding NATO's expansion toward its borders indicates a lingering, historical tension that will likely continue to be an issue for NATO in the future.

Extraterritorial Operations

THE PERIOD FOLLOWING the Cold War is marked by NATO involvement or action rather than sole reliance on passive deterrence. Deterrence remains an essential component of NATO's security strategy, but the alliance has taken a much more active approach to addressing security threats following the dissolution of the Soviet Union. Whereas NATO never responded to any crises or conflicts throughout the four decades of the Cold War, it has done so four times since then in a period half as long. While each of these conflicts occurred outside of NATO territory, NATO's involvement in Afghanistan was provoked by an attack on a member state.

This case appropriately reflects a more traditional definition of alliance response, although the targets and the conduct of the operation are very different from what would have been expected during the Cold War. The reasons (i.e., humanitarian concerns) for becoming involved in the other three conflicts were beyond any that NATO likely would have considered during the Cold War. This section explains why NATO became involved in each of the conflicts, the nature and successes of each response, and the broader implications of these efforts.

BOSNIA

NATO's FIRST EXTRATERRITORIAL mission was undertaken in 1995 in Bosnia for the purpose of peacekeeping. The reason for NATO's involvement in Bosnia was to end war crimes against the civilian population, which included ethnic cleansing, that erupted after Yugoslavia dissolved into a number of independent countries. A significant reason that led to NATO's involvement in resolving the Bosnian crisis was the failure of UN peacekeeping efforts, which caused the credibility of the UN to be questioned as the crisis continued.⁴⁶ Perhaps equally important, if not more important, was a lack of will among European states to deal with the crisis.⁴⁷

NATO's response to this crisis was consistent with its first post-Cold War Strategic Concept which indicated a desire to address events like this. Failure to respond would have implied an inability to adapt to the post-Soviet security environment.⁴⁸ The crisis in Bosnia had the potential for major, negative effects on the alliance. The delayed response demonstrated a lack of leadership among NATO allies and among the great powers in general. It also demonstrated weakness in European institutional (and NATO) ability to respond quickly to a major crisis, despite an expressed desire within Europe to take a stronger leadership role.⁴⁹ Furthermore, the UN was unable to quell the violence and prevent the ongoing genocide.⁵⁰

Although NATO participated in support of a UN-led operation to end hostilities in Bosnia, its role was limited to providing air support for UN troops, acting explicitly under the auspices of the UN.⁵¹ After playing a support role, NATO's involvement grew into an operation called Implementation Force (IFOR) with an initial deployment of 60,000 troops. IFOR's primary goal was to "maintain peace...[and] also, where necessary, to enforce it."⁵² It was able to meet each of its goals for the operation

within the time desired and immediately transferred its role over to the Stabilisation Force (SFOR).⁵³

“SFOR’s primary task was to contribute to a safe and secure environment conducive to civil and political reconstruction,” mainly by reducing elements that could facilitate “a resumption of hostilities” and by encouraging “confidence in the peace process.”⁵⁴ Its implementation occurred over the course of eight years and began with a military force of roughly half the size of IFOR’s.⁵⁵ SFOR ended in 2004 when the European Union (EU) became the primary organization overseeing civil development.⁵⁶

The sheer size of each of the operations in Bosnia, NATO’s inclusion of nonmembers and other intergovernmental institutions in the efforts, and the achievement of its goals are all indicators of NATO’s ability to successfully meet a major challenge presented by the new security environment that emerged in the wake of the Cold War. One of the more critical goals that was achieved by NATO was the operation’s one-year time line.⁵⁷ NATO’s success in addressing this post-Soviet threat was also illustrated by its extended presence beyond the immediate resolution of the conflict to help establish greater European stability. This contribution demonstrated that NATO was able to adjust from a deterrent-based posture, grounded in expectations of traditional military engagement, to one of a broader military and intensive civilian effort, characterized by NATO “as a peace support operation.”⁵⁸ The alliance’s long-term commitment to and evident success in creating a secure and stable environment helped bolster its credibility as a security-provider. Another positive result of the operations was that there was no spillover or greater regional instability. In addition, NATO operated extraterritorially in Bosnia. Although Bosnia is in Europe, it was not a member, nor a partner of the alliance at the time of the crisis. This was unprecedented. However, NATO’s failure to take an early leading role demonstrates a slow reaction to nontraditional missions, which ultimately contributed to greater numbers of civilian casualties, some of which could have been prevented through earlier involvement.

KOSOVO

NATO BECAME INVOLVED in another Balkans peacekeeping operation in Kosovo in 1999. Kosovo was a region contained within Serbia (both parts of the former Yugoslavia). The conflict in Kosovo arose because the Serbian government took brutal means to oppose

attempts at asserting Kosovo's independence from it. Like the one in Bosnia, this operation revealed much apprehension and limited leadership among NATO allies. NATO faced greater difficulty with intervening in Kosovo because of the issue of sovereignty: the UN had not sanctioned NATO's intervention in Kosovo as it had for the Bosnia crisis. As stated by Stephen J. Cimbala and Peter K. Forster, "Kosovo was a recognized part of Yugoslavia," and intervening there had the potential to establish a new norm that "human rights violations [would] become a *raison d'être* for violating national sovereignty," a concern for Russia and China, as both had been accused of such violations.⁵⁹ NATO made a number of threats of aerial bombardment against the Serbs in order to push them to the negotiation table; however, all such efforts yielded limited results. The two sides of the conflict were unable to agree on terms for peace, and ultimately it was a Serbian refusal which led to "[t]he 78-day air campaign."⁶⁰ Immediately following this campaign was the introduction of the Kosovo Force (KFOR) to quell ethnic violence, remove elements which could facilitate resurgent violence, and improve security, all while attempting to establish stable governance and provide "support to the international humanitarian effort."⁶¹

One factor that distinguishes the crisis in Kosovo from that in Bosnia is that the original operation remains active nearly thirteen years later. NATO has undertaken a number of major troop withdrawals over time in response to improving stability in the area. At one point, NATO had "progressively reduce[d] its presence to around 5,000 troops in total, marking one more step in the adaptation of KFOR to a deterrent presence," although the number is now slightly higher.⁶² However, the operation has not been entirely successful, as the root causes of the conflict have yet to be resolved. A resurgence of violence occurred in 2004, and the ensuing negotiations "failed to reach any agreement on Kosovo's status."⁶³

Despite a number of problems, this ongoing operation has produced some favorable results thus far, including ongoing improvement in the security of and stabilization of Kosovo (which were its primary goals). The fact that NATO became engaged in another extraterritorial mission, and remains so, is also significant. This is particularly the case if one considers the fact that KFOR was a response to yet another humanitarian crisis rather than a response to a traditional threat to the alliance. An area that requires much attention is the negotiation process because it has made no prog-

ress on resolving the root causes of the conflict.⁶⁴ As the operation has involved a number of troops from nonmembers and cooperation with the UN, Organization for Security and Cooperation in Europe (OSCE), and the EU, this conflict illustrates another case in which NATO demonstrated the capacity to make a decisive impact on the course of events in a crisis situation with actors beyond its member states and partners.⁶⁵

AFGHANISTAN

THE CONFLICT IN Afghanistan contains a number of features of the post-Cold War security environment that are starkly different from those of the Balkans crises. Involvement in Afghanistan has forced NATO to endure major growing pains as it struggles to adapt to meeting the challenges posed by this conflict. The first characteristic which makes NATO involvement in Afghanistan unlike either of the previous cases is that it was the first and only case in which Article V of the NATO Charter was invoked, the article which states the conditions for collective defense.⁶⁶ The second characteristic that distinguishes the current conflict from the Balkans crises is that the United States initially preferred not to operate under NATO auspices, even though Article V was invoked.⁶⁷ The third characteristic to consider is that NATO's involvement in Afghanistan has been an experience of ongoing hostilities that have not ended despite NATO's leadership role in a major military operation in the country (International Security Assistance Force – ISAF). Although instability has continued to be an issue in Kosovo, ongoing violence has not been a persistent feature in either Balkans conflict. In addition, in contrast to these crises, the level of troop deployment in Afghanistan has increased over time rather than decreased.

The intervention in Afghanistan was a reaction to an attack on a member state rather than a humanitarian crisis.⁶⁸ The attacks demonstrated that a nonstate entity could possess the capacity to directly attack one of NATO's allies, and thereby threaten the rest indirectly, whereas the Balkans crises did not pose such a clearly-defined, physical threat to alliance security.⁶⁹ The nature of the attacks, which consisted of hijacked airplanes being used as missiles, was also unprecedented and unanticipated. Furthermore, the conflict that has ensued in Afghanistan has required a major counterinsurgency effort by NATO and others participating, a very different type of conflict

than NATO anticipated having during the Cold War.⁷⁰

Canada reacted to the September 11, 2001 attacks by suggesting that NATO invoke Article V of the Charter.⁷¹ Each of NATO's members agreed, although the United States initially decided to act outside of NATO auspices.⁷² The Article V invocation should be regarded as a demonstration of alliance cohesion, yet NATO found itself marginalized by its largest member. However, the United States realized in 2003 that its stance was unsustainable because of the magnitude of the effort required. Thus, NATO assumed control of ISAF.⁷³

While each of the Balkans conflicts eventually helped bolster NATO's credibility, the slow progress in Afghanistan does not seem to have hurt it. Nonetheless, the ongoing hostilities in Afghanistan have challenged the potential for a successful resolution. Neither the American-led operation, nor the NATO-led one has successfully removed al-Qaeda or the Taliban from Afghanistan. Despite cautious progress with regard to establishing a stable democracy and security force, much remains to do before either task can be considered successful. Although there have been relatively few casualties for the alliance in Afghanistan, the financial cost of the conflict has been staggering.⁷⁴ In addition, "60 percent of Americans said they no longer believe the war in Afghanistan is worth fighting," as reported in December 2010.⁷⁵ The goals of the Afghanistan operation have been somewhat adjusted to reflect a different approach to the conflict, namely to couple successful counterinsurgency efforts "with good governance and economic opportunity as well as with improved security."⁷⁶ However, progress toward these ends has been slow and costly in manpower, resources, and financial expenditures. A positive, largely symbolic step was taken on May 1, 2011, when President Obama announced that after nearly ten years of conflict, Osama Bin Laden, the titular head of al-Qaeda, was finally killed in a military operation.⁷⁷ Despite the slow progress that had been made with regard to Bin Laden, the cost of the conflict, and the trouble establishing stable governance, public support exists nonetheless for the prospect of future extraterritorial operations (see Table 1). This support suggests that NATO continues to be viewed as a worthwhile security organization.

There has been a demonstrated need and push for a broad civilian effort to increase stability and governance, but such efforts can impact the complex intelligence aspects of the conflict in Afghanistan.⁷⁸ In this case, human intelligence plays a major role because the

primary adversaries are not regular nor identifiable army or militia, but targets mixed in with the population.⁷⁹ Fostering confidence in alliance forces could discourage sympathy for the Taliban and al-Qaeda because alliance forces would be seen as an alternative to these groups. This is a particular concern once one considers that the opium trade plays a major role in supporting both the Afghan population and the insurgency.⁸⁰ Public diplomacy efforts have been challenged, as many civilians perceive issues like diseases that have hurt opium crops as having NATO origins. Taliban propaganda has enhanced such perceptions and has attempted to connect these diseases with previous alliance attempts to destroy the crop.⁸¹

The conflicts in Bosnia, Kosovo, and Afghanistan have demonstrated that the security environment has changed since the Cold War and that NATO has adapted to this environment with cautious success. Despite a number of growing pains in NATO's adaptation, assertions of failure with respect to ISAF are premature. Although NATO has been in charge of a major operation in the conflict for nearly nine years, the conflict is projected to end after eleven years of its involvement. The leaders' willingness to agree to a cessation of NATO's combat presence in the country, a timetable set by Afghan President Hamid Karzai, is an indication of their confidence in achieving success.

This conflict is the only one in which Article V was invoked, and yet it has been the most challenging for the alliance, indicating that NATO was more successful in adapting to crisis management strategies than to counterinsurgency strategies. However, institutional flexibility is a slow process, as James G. March and Johan P. Olsen argue:

*Unless an environment is perfectly stable, or an institution instantaneously adaptive, of course, there will always be some delay in an adaptive process, thus some degree of mismatch between an environment and the institutions existing in it. But where an environment changes quickly relative to the rate at which an institution adapts, an adaptive process can easily and persistently fail to reach an equilibrium.*⁸²

Finally, NATO must consider the impact that this conflict will have on reducing its defense capabilities. The United States is expected to reduce the number of total active duty troops beginning in 2015.⁸³ This decision indicates the impact that both this conflict and the

conflict in Iraq have had on NATO's largest troop contributor. The cost of this U.S. troop reduction could be major, as European defense budgets have also been reduced in recent years.⁸⁴ The severity of these budget cuts casts a dark shadow on the future of the alliance and its ability to respond to future crises, as will be discussed later.

LIBYA

NATO'S INVOLVEMENT IN the recent conflict in Libya was quite different from the other conflicts discussed here. First, it was the only involvement in which the alliance did not contribute ground troops. Second, NATO's military involvement in the country ceased after only seven months. Third, it was the first operation in which NATO witnessed strong European leadership, with the United States accepting a less prominent leadership role than has ever been the case.

Western involvement in Libya began in response to civil unrest in the country, a result of Head of State Muammar Qaddafi using ruthless means to quell protests during the 2011 Arab Spring. European leaders took an early political position in support of Libyan rebels and sought to mobilize NATO as the primary enforcer of the UN resolution "to protect civilians in Libya."⁸⁵ The actions authorized in the resolution were, according to NATO, to enforce "an arms embargo, a no-fly zone [for non-humanitarian flights] and actions to protect civilians from attack or the threat of attack."⁸⁶ There was some initial dissent among NATO members (namely Germany and Turkey); however, the alliance eventually mobilized to enforce the resolution, taking over the role of the previous Paris-London-Washington coalition that had responded earlier to the call for an arms embargo.⁸⁷ NATO's official assumption of command began March 31, 2011, its operation consisting of both airborne and seaborne elements to enforce all aspects of the UN resolution. By the end of NATO involvement, the alliance had destroyed nearly six thousand targets and boarded roughly three hundred sea vessels throughout the conflict, as reported on October 24, 2011.⁸⁸ The operation officially ended on October 31, 2011 when the United Nations Security Council (UNSC) voted to end it in the wake of Qaddafi's death, his government's collapse, and the apparent military victory of the rebels.⁸⁹

One significant aspect of the operation is that it was successfully completed in seven months. In addition, the fact that NATO's role

was kept limited, but successful nonetheless, is important, as previous alliance interventions in humanitarian conflicts eventually led to the deployment of ground troops. Furthermore, the decision-making process to become involved in the Libyan conflict was also relatively quick, particularly given that it involved a response to a humanitarian crisis. Finally, NATO's ability to enlist the blessing of the Arab League was no small feat. This was a crucial development for the alliance not only because of its emphasis on cooperation with other international bodies, but because a number of countries within the Arab League faced domestic pressures similar to those in Libya, and thus had much at stake. That so many countries were willing to entrust NATO with the authority of enforcing the UN resolution points to its continued relevance in the international system.

Reducing the Nuclear Threat: Nuclear Arms Reduction and Nuclear Nonproliferation

ONE OF THE greatest contributions to NATO members' security in the post-Cold War period has been nuclear arms reduction. The reasons for this are multifold. Reducing nuclear arsenals increases predictability and confidence between former adversaries, which in turn reduces insecurity. Nuclear arms reduction also allows for a decreased likelihood of nuclear accidents. Finally, it lessens the vulnerability of nuclear material, as there are fewer targets for potential nuclear thieves. Efforts at reduction and control have been in conjunction with Russia. This is important as NATO and Russia are the two most heavily nuclear-armed entities in the world, so reductions in their arsenals will have the greatest impact on global nuclear arms reduction.

Before the official end of the Cold War, the first push for nuclear arms reduction between the United States and the Soviet Union occurred in 1991 with the Strategic Arms Reduction Treaty (START).⁹⁰ This treaty "called for a one-third reduction in the number of nuclear warheads and bombs held by" each party.⁹¹ This treaty also included former Soviet states which hosted Soviet-era weapons included in the treaty.⁹² Their inclusion was an important step in the nuclear weapons issue-area because it was successful and was the first of its kind.

American nuclear stockpiles and American-Russian nuclear arms reduction treaties are discussed here because, although both the French and British nuclear arsenals play a part in supporting the alli-

ance's nuclear force, "[t]he supreme guarantee of the security of the Allies continues to be provided by the strategic nuclear forces of the alliance, particularly those of the United States."⁹³ Russian nuclear weapons are also of great concern because of lingering insecurity among NATO allies, particularly former members of the Warsaw Pact.⁹⁴ Since the United States and Russia were enemies during the Cold War and continue to mistrust each other, reductions in nuclear arsenals will only occur simultaneously or reciprocally.⁹⁵ Any major decisions regarding alliance nuclear capabilities require "appropriate consultations among Allies."⁹⁶ It is for these reasons that NATO has a major role in nuclear arms reduction.

NATO's post-Cold War nuclear policy has echoed the START reductions that immediately followed the end of the Cold War:

*In the new security environment, NATO has radically reduced its reliance on nuclear forces. Its strategy remains one of war prevention but it is no longer dominated by the possibility of nuclear escalation. Its nuclear forces are no longer targeted against any country, and the circumstances in which their use might have to be contemplated are considered to be extremely remote.*⁹⁷

This demonstrates a shift from a nuclear posture directed at a specific source of threat to one of general deterrence.⁹⁸ This signifies that NATO has placed greater emphasis on facing broader security concerns that cannot effectively be dealt with using nuclear weapons. The reduced prominence of nuclear weapons partially reflects the recognition that other, non-nuclear military capabilities are expected to provide adequate security.⁹⁹

In 2002, another nuclear arms reduction treaty, the Moscow Treaty, was adopted. Like its predecessors, it sought to further reduce "the total number of strategic weapons...[this time to] 2,200."¹⁰⁰ Interestingly, this treaty "contained no verification means," although it was most likely intended to complement, rather than replace, the START treaty.¹⁰¹

In 2010, the New START, as it is popularly called, was written with the purpose of further reducing American and Russian deployed nuclear weapons. Because of NATO's dependence on American nuclear capabilities, and the historically unfavorable relationship between NATO and Russia, American-Russian relations have revolved primarily around the security of NATO's members,

as the American arsenal protects NATO's member-states under extended deterrence (hence the term *supreme guarantee*). This is a practice described as "[t]he threat to retaliate against an adversary with nuclear weapons on behalf of a third-party."¹⁰² The treaty seeks simultaneous ratification and implementation. It also requires greater transparency, which has obvious positive implications in terms of relations between the United States (and NATO by extension) and Russia.¹⁰³ In addition, future Russian cooperation on Iranian sanctions and cooperation with the conflict in Afghanistan may have been in jeopardy had New START not been ratified.¹⁰⁴

New START was ratified by the United States Senate in December 2010.¹⁰⁵ However, the New START should not be viewed as a cure-all in terms of reducing the nuclear threat that NATO allies face. Peter Baker states that, first, it does not require "curbing tactical nuclear weapons," which pose a potential threat given that "Russia has far more such weapons than the United States...[and that] as recently as last spring Russia moved some of them closer to its borders with NATO nations as a n [*sic*] response to American missile-defense deployments."¹⁰⁶ Second, "[s]ome experts consider these smaller bombs a greater risk of theft or black-market diversion to rogue states or terrorist groups."¹⁰⁷ This introduces the issue of nuclear proliferation, which may in fact pose a greater nuclear threat to the alliance than deployed Russian strategic warheads (assuming that even these are secure). Russian nuclear facilities have been a source of concern for NATO throughout the post-Cold War period because of the vulnerability of existing nuclear technology and material.¹⁰⁸ This is a very real issue because the former Soviet Union had a large nuclear arsenal and cannot account for some of such material.¹⁰⁹ Illicit Russian nuclear proliferation has also occurred as a result of "underpaid nuclear scientists, and the smuggling of nuclear materials."¹¹⁰

All of the above present significant threats to alliance security because weaknesses in the security of nuclear materials can contribute to the spread of nuclear knowledge to states and non-state actors. Despite NATO's belief that its military capabilities are an effective deterrent, there is great concern about the potential for "irrational [actors] by our [the alliance's] standards" gaining nuclear capabilities.¹¹¹ Al-Qaeda's September 11, 2001 attacks on the United States present a case in which deterrence had little impact on preventing transnational groups from inflicting massive amounts

of damage at relatively low cost without the use of nuclear weapons. A nuclear-armed transnational terrorist group like al-Qaeda is clearly a major concern.

The amount of material lost and the number of scientists who have aided states and transnational actors in acquiring nuclear technology and knowledge are both unknown. Therefore, it is difficult to measure the success of NATO's nonproliferation efforts. Perhaps the most appropriate indicator at this point would simply be the absence of nuclear attacks on or within the alliance, at best suggesting that nonproliferation efforts have been effective thus far. Whether this has been a result of NATO's efforts, a combination of efforts including other actors, or simply the incapacity of other actors to successfully develop nuclear weapons is impossible to tell. Securing nuclear sites within NATO and Russia and reducing the number of nuclear arms offer a partial solution. The difficulty lies in successfully discouraging other actors from seeking nuclear technology.¹¹² Such an issue requires cooperation and transparency with a significant number of other states in the international system. To this end, NATO supports the Nuclear Non-Proliferation Treaty (NPT) and the International Atomic Energy Agency (IAEA) by extension, because the latter "verifies through its inspection system that states comply with their commitments, under the Non-Proliferation Treaty and other non-proliferation agreements, to use nuclear material and facilities only for peaceful purposes."¹¹³ Cooperation with intergovernmental organizations increases transparency and reduces uncertainty in the international system, especially if it can be demonstrated that states deliver on their commitments. A problem with this approach is that some states have not complied with the treaty and have even withdrawn from it (i.e., Iran and North Korea, respectively). NATO itself is not fully compliant with the NPT on a number of issues. First, NATO's member states have not eliminated their nuclear stockpiles.¹¹⁴ Second, members of NATO have shared nuclear technology and materials with other states. In one case, in American cooperation with Pakistan on the latter's nuclear development, the beneficiary had weak security over its stockpiles.¹¹⁵

Committing to the NPT and IAEA demonstrates that NATO seeks means beyond its capabilities to reduce nuclear proliferation, although for the reasons stated above, this may not be the most effective approach. A double-standard complicates the matter, as NATO will always seek to maintain the security of its member states by any means it deems necessary. Nevertheless, NATO has made significant

progress toward improving nuclear security, as mentioned above. Although NATO maintains a nuclear arsenal for security reasons, a lack of adherence to the treaty by any actor contributes to insecurity because it enhances unpredictability in the international system. Even isolated nuclear proliferation is a concern for NATO. The destructive power of only one nuclear weapon is of concern enough, let alone multiple nuclear weapons under the control of multiple, unpredictable actors. Further, a lack of adherence to the NPT undermines the treaty's legitimacy, and thus could discourage adherence by other states. This demonstrates that nonproliferation requires discouraging entities from pursuing nuclear capabilities, as well as dissuading the desire for others to provide such capabilities.¹¹⁶

Iran is an actor that has been a particular concern for NATO. Iran has not been transparent in terms of its nuclear pursuits and has had limited compliance with NPT.¹¹⁷ Iran continues to enrich uranium in spite of years of sanctions, which could demonstrate a significant weakness of nonproliferation efforts, regardless of previous successes.¹¹⁸ A nuclear Iran also has the potential to sponsor misuse or further nuclear proliferation in the region.¹¹⁹

The case of Iran also demonstrates a way in which deterrence has had little impact, at least in terms of its pursuit of nuclear technology (Iran still publicly maintains that it is seeking nuclear technology for nonmilitary uses).¹²⁰ Although the actions of a nuclear Iran would be somewhat unpredictable, NATO expects that its own current military capabilities should be enough of a deterrent to discourage any form of attack on alliance members.¹²¹

NATO has had mixed success in reducing nuclear stockpiles and limiting nuclear proliferation. Although it has made significant progress in reducing some types of weapons in both NATO member states' and Russia's nuclear arsenals, more vulnerable types have not been sufficiently reduced or eliminated. Reducing stockpiles has also had an indirect impact on curbing nuclear proliferation because it has reduced the amount of nuclear technology that could be stolen by terrorists or sold by disgruntled scientists. The EU's expressed desire to contribute to nuclear nonproliferation will give NATO another organization with which it can cooperate to reduce the nuclear threat. NATO's current cooperation with the NPT and the IAEA further demonstrates a desire to adjust to the post-Cold War security environment.

IDENTITY

ROLE AND FUNCTION offer a concrete foundation for explaining why NATO has continued to be relevant throughout the post-Cold War period. The functions detailed above demonstrate ways in which NATO has had a major impact on the European security environment in a post-Soviet context. A major reason why NATO continues to respond to major security issues is that each of its member states believes that NATO offers the most security at essentially no cost to them (with some exceptions). This expectation points to a deeper, underlying feature of the alliance that moves beyond simple rational calculation. Although the alliance was formed as a reflection of rational state behavior in reaction to the Cold War, an alliance-based identity grew as time passed.¹²² The most useful theoretical approach to explain this phenomenon is the concept of community identity. According to Amitav Acharya, community identities can arise from a variety of sources:

*Constructivists argue that collective identities among states are constructed by their social interactions, rather than given exogenously to them by human nature, domestic politics, or, one might add, the international distribution of power. Viewed in these terms, regional cooperation among states is not necessarily a function of immutable or pre-ordained variables such as physical location, common historical experience, level of economic development, shared values, cultural affinities, and ideological convergence. Rather, regionalism may emerge and consolidate itself within an intersubjective setting of dynamic interactions consisting of shared understandings, expectations, and social knowledge embedded in international institutions and threat complexes...*¹²³

Building on this concept, one can consider the idea of communities of practice. Communities of practice are the products of actors' behaviors, which become the main feature of the group. Practices become, more or less, established means of interaction within the group.¹²⁴

A specific form of a community of practice is a security community, in which each member of a group of actors has "dependable expectations of peaceful change."¹²⁵ A major component of this, a term Emmanuel Adler calls "self-restraint," is best understood

as mutual benevolence among members of the security community. All actors engage in non-confrontational relations. In order for this to be a concrete feature of security communities, the common practice of self-restraint must become an institutionalized feature of communities.¹²⁶

The institutionalization of self-restraint has indeed been a feature of NATO since its inception.¹²⁷ The emphasis on constant consultation among member states contributes to self-restraint because it provides greater cohesion and transparency among the member states.¹²⁸ Both self-restraint and consultation are parts of a broader concept of reciprocity, which implies a shared vision of the future.¹²⁹ The absence of self-restraint would discourage nonviolent relations, and, by definition, a security community could not exist. With self-restraint a long-established practice, member states have demonstrated a greater amount of trust in the alliance. This trust is also evident in intelligence-sharing, for example.¹³⁰ Interestingly, “Europe remains heavily dependent on U.S. intelligence leaders,” which is additional evidence that the level of trust in alliance members is significant.¹³¹ This is the case because intelligence-sharing, by definition, requires the sharing of incredibly delicate information that could increase each state’s vulnerability vis-à-vis each other (not unlike the Prisoner’s Dilemma).

Common practices can lead to the emergence of regionalism, a very relevant aspect of NATO’s identity. Regionalism is regarded here as an image of geographic spatiality and a psychological association with that geographic entity. The extended quote above states that a regional identity is developed more out of interactions than geography. This notion is supported by some who posit that NATO and European identity had to grow metaphysically before the alliance could physically enlarge.¹³² This was important because even the name of the alliance implies a specific geographic orientation, namely the “Euro-Atlantic area,” which can be conflated with a *Western* identity.¹³³ With a geographically-related identity came a variety of less-concrete political and social associations, with greater implications than spatial relations.¹³⁴

NATO’s association with the West is important because of a historical divide between East and West. Generally speaking, this divide has been a product of political, social, and economic stereotypes that the West has associated with the East, such as the latter having a particularly undeveloped nature about it. These stereotypes were typically cast as the antithesis of the features that character-

ized the West, allowing the West an *other* against which it could define itself, a process which appears to have reinforced such sentiments.¹³⁵ This notion, a manifestation of Orientalism-like behavior occurring within Europe, demonstrates pervasive Western patronizing of the East, where “the idea of European identity [i]s a superior one in comparison with all the non-European peoples and cultures.”¹³⁶ Since the Eastern half of Europe is considered separate from the *European* identity, observers understand the extent to which this identity derives from a social construct rather than from purely geographical origins.

For expansion of the NATO identity to occur, and perhaps more broadly, the Western identity, the possibility of identity expansion needed to be considered.¹³⁷ This clearly occurred as “conflict in Southeastern Europe had been redefined as conflict in *Europe*” even before the demise of the Soviet Union.¹³⁸ Because NATO had more or less presented itself “as [the] guarantor of Western civilization” when it took action in Bosnia, the metaphysical expansion of the alliance’s identity became more concrete.¹³⁹ The fact that this was able to occur in such a short period following centuries of division and decades of overt, ideological tension should not be ignored.

It is important to recognize, however, that the alliance had to act in order to protect the Western identity’s historical commitment to liberal ideals. Had it not acted in Bosnia, NATO’s credibility would have been called into question.¹⁴⁰ It is also important to recognize the distinction between the *West* and *NATO*. First, if the credibility of the West was actually at stake, NATO’s involvement in Bosnia was to protect the Western liberal credentials.¹⁴¹ Second, NATO is the enforcement mechanism of the West, despite the fact that its membership does not include all countries associated with the West. Its inability to protect the West would have called into question its own credibility as a viable security organization. Finally, the process of identity expansion, as compared to physical enlargement, appears to be much more difficult.

Ultimately, expansion occurs on the basis of sharing common values and common practices. Although these requirements reflect common values, they also reflect common practices.¹⁴² Adherence to both is essential, as the alliance itself evolved to include both common practices and values as fundamental features. If existing members are expected to adhere to these practices and values, then new members should as well (with the exception of Albania, as discussed earlier). Common practices were addressed before common

values because an emphasis on the former may be regarded as a precursor to the growth of the latter.¹⁴³ The practice of contributing to collective security was initially important, but greater emphasis has been placed on the necessity of democratic governance among member states since the end of the Cold War.

With time and the deterioration of a clear military threat, the significance of common values has grown. Perhaps the most striking value that the alliance has asserted as a necessary feature for new members is an established (although somewhat loosely-defined) democratic regime. Although alliance identity expanded to include former adversaries (states that represented a historically-stigmatized *other*), the democratic prerequisite suggests that NATO sought to adjust these actors to fit a more Western identity. With that said, the identity of the West has not changed much. What has changed is the geographic size of the area referred to as the West.

The theoretical aspect implies that the Western identity has been conflated with the NATO identity. Indeed, there has been some encouraging data to support this. A recent public opinion poll that surveyed citizens in thirteen NATO member states found that a majority of those surveyed agreed with the notion that NATO is an essential component of their respective country's security (see Table 2). These findings have important implications. The findings imply a widely-shared acceptance of NATO's role in providing security to its members. The large number of supporters for this notion suggests that the interviewed citizens view themselves as part of the West. In a counter-example, as Turkey and the West have been growing more distant in the past several years, the plurality of negative responses to this question in Turkey suggests that it identifies less with the West than do other member states.¹⁴⁴

The fact that so many agree that NATO is essential to their respective countries' security implies a widely-shared sense of community. A shared sense of destiny is implicit, as a significant number believe that NATO will protect their security. In addition, self-restraint, being the predominant practice of the community, has become an expectation. It is unlikely that individuals would regard NATO as an essential security entity if it was not expected to maintain non-confrontational relations among its member states.

Although the survey results indicate belief in NATO relevance in the post-Cold War period, one must be cautious about the strength of this identity. Persistent, disproportionate burden-sharing

poses a challenge to identity strength, although identity cannot be reduced solely to defense spending. Financial commitments do not necessarily weaken NATO's identity. They do elucidate, however, the link between unrivaled American leadership within NATO and its disproportionate share of bankrolling the alliance.

Intergovernmental burden-sharing can be measured in terms of spending allotted for a particular purpose as a percentage of each member state's gross domestic product (GDP).¹⁴⁵ In NATO's case, burden-sharing is measured as a percentage of GDP allotted for defense.¹⁴⁶ Throughout NATO's existence, the United States has borne a disproportionate share of the burden to support the alliance. In the organization's early years, unequal burden-sharing could be attributed to the damage caused by WWII. The United States' allies were devastated by the destruction they had experienced throughout the war, while the United States emerged from the war with a strong economy and a preponderant military capability. However, this uneven result does not explain why this pattern continued into the present. There have been a few exceptions where other members have borne higher burdens, although these should be regarded as isolated cases.¹⁴⁷

The function of an alliance has an impact on burden-sharing; alliances in which deterrence is the primary concern open the possibility of free-riding. However, when actual defense is at stake for individual members, they will be more likely to contribute to the alliance.¹⁴⁸ This would appear to be relevant today, both because of the absence of a clear source of threat to the alliance, and because so many members spend less on defense as a percentage of GDP as compared to the United States.¹⁴⁹

There has been a significantly unequal distribution of burden-sharing among NATO members between the years 2005 and 2009. NATO members had agreed to meet a minimum threshold of 2% of GDP for defense spending, but only eight member states met or exceeded that level during this time frame.¹⁵⁰ A decrease in defense spending could imply a significant decrease in traditional security concerns. Because NATO does not identify a particular state as a source of threat, it provides general deterrence against any potential aggressors.¹⁵¹ Thus, there is a lower incentive for smaller members to contribute to collective security.¹⁵² Although there is an apparent unwillingness to meet the minimum threshold for defense spending, NATO identity has not weakened. NATO member states trust the United States' commitment to the alliance enough to allow

American leadership in times of crisis. The increasingly broad range of security concerns in the post-Cold War era include a number of nontraditional security concerns that traditional defense cannot address. These points are interrelated. The United States provides the supreme guarantee to the alliance via nuclear weapons at essentially no cost to the other member states (except perhaps for some political costs to those hosting the weapons). Despite the Article V invocation which requires a collective response to an attack on any member, the United States initially chose to operate outside of NATO on Afghanistan. The United States still provides more than two-thirds of the troops participating in this operation (ISAF).¹⁵³ Though this indicates marginalization of the alliance by its strongest member, the alliance indicated that it was willing to operate collectively by invoking the collective defense article. In addition, NATO member states enjoy the benefits of enlargement, nuclear arms reduction and nonproliferation, and peacekeeping operations, all at a low cost of membership.

As discussed earlier, the operation in Libya demonstrated stronger European leadership, but did not resolve the issue of burden-sharing. Throughout the Cold War, burden-sharing was generally treated as a non-issue.¹⁵⁴ However, non-American member states continuing to bear little of the defense burden raises concerns because the United States has sought a smaller public leadership role and expressed the desire for a smaller operational role in Libya (despite the fact that it expanded its role during the course of the operation).¹⁵⁵ Future operations in which the U.S. may continue this trend, but with more concrete manifestations (such as fewer troop and equipment commitments), could bode ill for the alliance unless European members are more creative with their approach to supporting security initiatives (in the absence of increasing defense expenditures).¹⁵⁶

Although there have been concerns among some member states about the American commitment to protecting the alliance, the case of Libya presents the opposite, in which the United States is very much involved in the operation with a somewhat diminished leadership role.¹⁵⁷ The United States' willingness to bear a disproportionate financial burden for this crisis is a deviation from the norm because of this leadership adjustment. Mark Webber captured the historical concern of American leadership in the alliance when he wrote that "[t]he exceptional position of the U.S. within the alliance, coupled with its standing as a global power, has meant that Ameri-

can foreign policy, while often made by reference to NATO, has never been subordinate to it.”¹⁵⁸

The recent decline in defense spending among European member states has occurred in a time of recession. This, coupled with an aversion to the human cost of conflict, has encouraged lower defense spending among many European NATO members. Britain, the second largest troop contributor to the Afghanistan conflict, is expected to reduce defense spending by eight percent over the next three years.¹⁵⁹ Although the spending reduction was described as an active way to improve the cost-effectiveness of the military, diverging opinions over how defense money should be spent, particularly in terms of how European states should adapt to the current security environment, have been prevalent.¹⁶⁰ The United States is also grappling with an urgent pressure to reduce government spending, which might include the defense budget.¹⁶¹ This brings to light the shift that has occurred in security thinking in the post-Cold War period. Issues beyond military security are of a greater concern than they were during the Cold War.¹⁶² In the middle of the conflict in Libya, the Secretary General of NATO had already begun to show trepidation with regard to the strength of the alliance if future military spending continues along its pattern since the end of the Cold War.¹⁶³

A push for a greater European-based defense identity, be it within the EU or within NATO, is an additional challenge to NATO. Hints and manifestations of attempts at forging a European security identity have been a part of the European security dialogue since NATO's establishment.¹⁶⁴ Though NATO has clearly been the predominant security organization for Europe (and the West, more broadly) since its inception, this has not discouraged attempts to assert a stronger European identity by some member states.¹⁶⁵ France, separate from the integrated military structure for much of the alliance's existence, has been perhaps the staunchest supporter of a stronger European defense identity.¹⁶⁶ It is most likely the case that France rejoined the command structure with the intention of exerting greater influence on the Western security structure. At times, European security identities have largely been in support of NATO, if not housed within NATO itself, generally with the purpose of responding to immediate security and humanitarian crises rather than collective defense.¹⁶⁷ Although there has been enthusiasm among some European states, substantial progress has not been made in this endeavor. Bastian

Giegerich cited a tendency for European states to continue to deal with defense issues on a national level with little regard for the broader implications.¹⁶⁸ Furthermore, it would appear that there is a lack of willingness to meet commitments to ambitious initiatives that would create a more concrete face for a European-based security identity.¹⁶⁹ Even attempts at forging a common European foreign policy housed within the EU have been incredibly difficult, as the EU foreign policy chief lacks the support among the Union's member states to build a foreign policy.¹⁷⁰

Another issue that challenges the development of European-based security organizations (which would bolster a European security identity) is similar to what hampers European support for NATO, specifically defense spending. The cost of creating a new European collective organization, or significantly altering an existing one, would come at a great cost.¹⁷¹ Although European member states sought a greater leadership role with the operation in Libya, this was a difficult effort, even without the possibility of troop deployment in Libya. If this crisis elevated to a level in which troop deployment was considered, European leadership may have been seriously hampered.

Although a NATO identity clearly exists in Europe, it has faced significant challenges. These challenges stem from low defense spending and sporadic attempts at forging a European-based security identity, either within NATO or within another intergovernmental institution (including the EU). The root cause of this inability to create a strong European identity is related to the issue of leadership, as the United States fills this role and accepts this position although there are disparities in defense spending. Despite failures in European burden-sharing and a persistent dominant leadership role for the United States, Europeans have recently demonstrated a willingness to take the initiative to respond to an extraterritorial crisis and emphasized the necessity of NATO for orchestrating this response. Because of this renewed enthusiasm for the alliance and the existence of public opinion data indicating popular support for it, it is possible to conclude that NATO maintains a strong identity.

NATO'S POST-SOVIET RELEVANCE

NATO WAS SUCCESSFUL during the Cold War because it was able to deter Soviet aggression. However, the dissolution of its pri-

mary purpose did not render the alliance irrelevant. On the contrary, the alliance has enlarged to include a number of Cold War adversaries, involved itself in four extraterritorial conflicts, and contributed to nuclear arms reduction, all in a period half as long as the Cold War. The fact that the majority of NATO's extraterritorial operations have been nontraditional, humanitarian actions is a major departure from Cold War-era behavior. Early in the post-Soviet period, NATO's attempt to adapt was not as rapid as some would have liked; leadership reluctance hindered responses to emerging security threats. The alliance appears to have learned from this and has met each successive crisis with greater effectiveness. However, the problem of European defense cuts — and possible American ones — will continue to color the future of the alliance.¹⁷² In the midst of the Libyan operation, Anders Fogh Rasmussen, the Secretary General of NATO, expressed pessimism because of decreases in European defense budgets: "At the current pace of [defense budget] cuts, it is hard to see how Europe could maintain enough military capabilities to sustain similar operations in the future."¹⁷³ Robert Gates echoed a similar perspective in his final budget testimony to the U.S. Congress as the U.S. Secretary of Defense. He stated that the low levels of defense spending among European member states will be problematic, as future American leaders may be less likely to continue to bear such a disproportionate burden.¹⁷⁴ Rasmussen also said, "Any shortfalls have been primarily due to political, rather than military, constraints."¹⁷⁵ If political and fiscal issues continue to impact Europe's military capabilities to such a great extent, then the type of constraints may not matter — the damage to military infrastructure will already be done. Ultimately, it is likely that NATO will show the will to respond to future crises like the one it faced in Libya. What should be expected, however, is that the military response will likely be more dependent on the Americans than ever. Another factor to consider with regard to future security arrangements is the nature of security itself. As has been demonstrated above, economic concerns have become increasingly important with regard to security-provision. NATO may not be suited to deal with some aspects of security, including environmental disaster response. Also, given economic constraints on the military, the alliance may pursue more strategic operations in the future, rather than humanitarian ones.

Identity continues to be a relevant feature of the alliance today. Despite the notion that the alliance is seen by some as a cover for American foreign policy, it appears that other member states identify

strongly with NATO. Members enjoy the benefit of security at very low costs. As Europeans continue to seek a regional security identity, NATO will most likely be the most useful vehicle for such pursuits.

TABLE 1 The Extent to Which Future Extraterritorial Operations Would Be Supported

	US	FR	GER	UK	IT	NL	PL	PT	SP	SK	TR	BG	RO
Europe-Only, Article-5	21%	28%	41%	38%	25%	30%	30%	21%	27%	25%	17%	31%	37%
Beyond Europe	77%	68%	55%	59%	67%	67%	61%	76%	68%	61%	48%	45%	42%
Don't Know/Refused	1%	4%	3%	3%	8%	3%	9%	3%	5%	14%	34%	24%	21%

SOURCE: Topline Data, *Transatlantic Trends 2010*, German Marshall Fund of the United States, June 2010, Q12, Page 38. Responses to: "Some people say that NATO should limit its mission to defending members that are attacked in Europe. Others say that NATO must also be prepared to act outside of Europe to defend members from threats to their security. Which view is closer to your own?"

TABLE 2 The Extent to Which NATO is Regarded as an Essential Component of Each Surveyed Member States' Security

	US	FR	GER	UK	IT	NL	PL	PT	SP	SK	TR	BG	RO
Still Essential	60%	60%	56%	68%	54%	72%	52%	67%	57%	64%	30%	60%	65%
No Longer Essential	29%	34%	41%	29%	38%	26%	37%	31%	38%	22%	43%	23%	17%
Don't Know/Refused	11%	6%	3%	3%	9%	3%	11%	3%	4%	14%	27%	16%	18%

SOURCE: Topline Data, *Transatlantic Trends 2010*, German Marshall Fund of the United States, June 2010, Q11, Page 37. Responses to: "Some people say that NATO is essential to our country's security. Others say it is no longer essential. Which of these views is closer to your own?"

ENDNOTES

- 1 Robert McCalla, "NATO's Persistence after the Cold War," *International Organization* 50.3 (1996): 445-475. McCalla also argues that traditional theories do not account for NATO's post-Cold War success, although his thesis relies primarily upon theoretical grounds. Regardless, the present article does support his conclusion.
- 2 Kenneth N. Waltz, *Theory of International Politics* (New York: McGraw Hill, 1979) 106, 109, 118.
- 3 See Stephen Walt, "Alliance Formation and the Balance of World Power," *International Security* 9.4 (1985): 7-8.
- 4 Walt 4.
- 5 Walt 8, 17.
- 6 Edward Vose Gulick, *Europe's Classical Balance of Power* (New York: W.W. Norton and Company, 1967) 33-34.
- 7 Walt 8-9
- 8 See Walt 9-13
- 9 Waltz, *Theory* 126.
- 10 See Kenneth N. Waltz, "The Emerging Structure of International Politics," *International Security* 18.4 (1993); 44-79.
- 11 McCalla 448.
- 12 Alice Ba and Matthew J. Hoffman, "Making and Re-Making the World for IR 101: A Resource for Teaching Social Constructivism in Introductory Classes," *International Studies Perspectives* 4.1 (2003): 20.
- 13 Ted Hopf, "The Promise of Constructivism in International Relations Theory," *International Security* 23.1 (1998): 182.
- 14 Rey Koslowski and Friederich V. Kratochwil, "Understanding Change in International Politics: The Soviet Empire's Demise and the International System," *International Organization* 48.2 (1994): 222.
- 15 Robert O. Keohane, "The Demand for International Regimes," *International Organization* 36.2 (1982): 337-338, 342. Although NATO is not a regime, it could be considered the institutionalization of a Western security regime.
- 16 Celeste A. Wallander, "Institutional Assets and Adaptability: NATO After the Cold War," *International Organization* 54.4 (2000): 706-707; McCalla discusses this as well: 462-464.
- 17 See Waltz, *Theory* Chapter 6
- 18 McGeorge Bundy's *Danger and Survival* is an excellent resource on U.S. nuclear decision making throughout much of the Cold War.
- 19 See "The Alliance's New Strategic Concept," *NATO.int*. 7-8 Nov. 1991 (last updated 26 Aug 2010). North Atlantic Treaty Organization. 17 Nov. 2010 <http://www.nato.int/cps/en/natolive/official_texts_23847.htm>; See also Charles L. Glaser, "Why NATO is Still Best: Future Security Arrangements for Europe," *International Security* 18.1 (1993): 5-50.
- 20 "Strategic Concept For the Defence and Security of The Members of the North Atlantic Treaty Organisation," *NATO.int*. 19 Nov. 2010. North Atlantic Treaty Organization. 21 Nov. 2010 <<http://www.nato.int/lisbon2010/strategic-concept-2010-eng.pdf>>.
- 21 Chester A. Crocker, "Engaging Failing States," *Foreign Affairs* 82.5: 34.
- 22 Osama Bin Laden, al-Qaeda, and Afghanistan were quickly suspected in the immediate aftermath of September 11, 2001; Elisabeth Bumiller and Jane Perlez, "After the Attacks: The Overview; Bush and Top Aides Proclaim Policy of 'Ending' States that Back Terror; Local Airports Shut After an Arrest," *The New York Times* 14 Sept. 2001. 6 Dec. 2010 <<http://www.nytimes.com/2001/09/14/us/after-attacks-overview-bush-top-aides-proclaim-policy-ending-states-that-back.html?pagewanted=all>>.

- 23 See Thomas Schweich, "Is Afghanistan a Narco-State?" *The New York Times* 27 July 2008. 6 Dec. 2010 <<http://www.nytimes.com/2008/07/27/magazine/27AFGHAN-t.html>> and James Risen, "Poppy Fields are Now a Front Line in Afghan War," *The New York Times* 16 May 2007. 6 Dec 2010 <<http://www.nytimes.com/2007/05/16/world/asia/16drugs.html>>.
- 24 "Strategic Concept."
- 25 "NATO Enlargement," *NATO.int*. 23 Nov. 2010. North Atlantic Treaty Organization. 12 Jan. 2011 <http://www.nato.int/cps/en/natolive/topics_49212.htm>; "The Partnership for Peace Programme," *NATO.int*. 9 Nov. 2010. North Atlantic Treaty Organization. 12 Jan. 2011 <http://www.nato.int/cps/en/natolive/topics_50349.htm>.
- 26 Greece and Turkey joined in 1952, Germany in 1955, and Spain in 1982; "NATO Enlargement."
- 27 "Study on NATO Enlargement." *NATO.int*. Sept. 1995 (last updated 05 Nov. 2008). North Atlantic Treaty Organization. 16 Nov. 2011 <http://www.nato.int/cps/en/natolive/official_texts_24733.htm>.
- 28 "Study on NATO Enlargement."
- 29 "Study on NATO Enlargement."
- 30 See Maria Mälksoo, "Enabling NATO Enlargement: Changing Constructions of the Baltic States," *Trames* 8.3 (2004).
- 31 "The Partnership for Peace Programme."
- 32 See "Study on NATO Enlargement."
- 33 Helene Sjursen, "On the Identity of NATO," *International Affairs* 80.4 (2004): 695.
- 34 Sjursen 695.
- 35 "NATO Enlargement"
- 36 Bulgaria, Latvia, Estonia, Slovakia, Slovenia and Romania. "Map of Freedom in the World: Map of Freedom 2004," *Freedomhouse.org*. Freedom House, Inc. 24 Jan. 2011. <<http://www.freedomhouse.org/template.cfm?page=363&year=2004>>; Albania. "Map of Freedom in the World: Map of Freedom 2009," *Freedomhouse.org*. Freedom House, Inc. 24 Jan. 2011. <<http://www.freedomhouse.org/template.cfm?page=363&year=2009>>; Albania, "Map of Freedom in the World: Map of Freedom 2010," *Freedomhouse.org*. Freedom House, Inc. 24 Jan. 2011. <<http://www.freedomhouse.org/template.cfm?page=363&year=2010>>.
- 37 Sjursen 701.
- 38 Anne Applebaum, "A NATO for the 21st century," *The Washington Post* 23 Nov. 2010. 7 Dec. 2010 <<http://www.washingtonpost.com/wp-dyn/content/article/2010/11/22/AR2010112205558.html>>.
- 39 Dmitri K. Simes, "Losing Russia," *Foreign Affairs* 86.6 (2007): 44.
- 40 Dusan Stojanovic, "Serbia Shifting West, Away Russia," *The Associated Press* 23 Nov. 2010. 13 Jan. 2011 <<http://www.huffingtonpost.com/huff-wires/20101123/eu-serbia-looking-west/>>.
- 41 Simes 36.
- 42 Brian Frederking, "Constructing Post-Cold War Collective Security," *The American Political Science Review* 97.3 (2003): 371; Derek E. Mix, "The United States and Europe; Current Issues," *Congressional Research Service*. 20 June 2011. 21 Nov. 2011 <<http://www.fas.org/spp/crs/row/RS22163.pdf>>: 9.
- 43 "NATO-Russia Council," *NATO.int*. 27 Oct. 2010. North Atlantic Treaty Organization. 13 Jan. 2011 <http://www.nato.int/cps/en/natolive/topics_50091.htm>.
- 44 Mix 9.
- 45 Mix 9.
- 46 Ryan C. Hendrickson, *Diplomacy and War at NATO: The Secretary General and Military Action after the Cold War* (Columbia: University of Missouri Press, 2006) 49-55; Stephen J. Cimbala and Peter K. Forster, *Multinational Military Intervention:*

NATO Policy, Strategy, and Burden Sharing (Burlington: Ashgate, 2010) 100, 101, and see 97-148.

- 47 Hendrickson 49-55.
- 48 Evident by the date of “The Alliance’s New Strategic Concept”; “Peace Support Operations in Bosnia and Herzegovina,” *NATO.int*. 25 Oct. 2010. North Atlantic Treaty Organization. 13 Jan. 2011 <http://www.nato.int/cps/en/natolive/top-ics_52122.htm>.
- 49 Cimbala and Forster 97-104.
- 50 Dan Caldwell and Robert E. Williams, Jr., *Seeking Security in an Insecure World* (Lanham, MD: Rowman and Littlefield Publishers, Inc. 2006) 135-136.
- 51 Cimbala and Forster 103; “Peace Support Operations in Bosnia and Herzegovina.”
- 52 “Peace Support Operations in Bosnia and Herzegovina.”
- 53 “History of the NATO-led Stabilization Force (SFOR) in Bosnia and Herzegovina.” *SFOR:Stabilization Force*. North Atlantic Treaty Organization. 24 Nov. 2010 <<http://www.nato.int/sfor/docu/d981116a.htm>>.
- 54 “Peace Support Operations in Bosnia and Herzegovina.”
- 55 “History of the NATO-led Stabilization Force (SFOR) in Bosnia and Herzegovina.”
- 56 “NATO’s relations with Bosnia and Herzegovina,” *NATO.int*. 10 Nov. 2010. North Atlantic Treaty Organization. 11 Jan. 2011 <http://www.nato.int/cps/en/natolive/topics_49127.htm>.
- 57 Cimbala and Forster 101, 103, 135-136; “History of the NATO-led Stabilization Forces (SFOR) in Bosnia and Herzegovina.”
- 58 “Peace Support Operations in Bosnia and Herzegovina.”
- 59 Cimbala and Forster 116.
- 60 Cimbala and Forster 116.
- 61 “NATO’s role in Kosovo,” *NATO.int*. 10 Nov. 2010. North Atlantic Treaty Organization. 23 Nov. 2011 <http://www.nato.int/cps/en/natolive/topics_48818.htm>.
- 62 “Statement by the NATO Secretary General: Progress in Kosovo Allows for Adaptation of KFOR Presence,” *NATO.int*. 29 Oct. 2010. North Atlantic Treaty Organization. 11 Jan. 2011 <http://www.nato.int/cps/en/natolive/news_67610.htm?selectedLocale=en>; According to “NATO’s role in Kosovo,” the number is now 5,500 troops.
- 63 “NATO’s role in Kosovo.”
- 64 “Defence Ministers Agree: KFOR to Remain as Long as Necessary,” *NATO.int*. 10 June 2010. North Atlantic Treaty Organization. 11 Jan. 2011. <http://www.nato.int/cps/en/natolive/news_64266.htm?selectedLocale=en>; “NATO’s role in Kosovo.”
- 65 Cimbala and Forster 116.
- 66 Hendrickson 120; Greater details about the nature of this particular Article can be found in “Collective Defence,” *NATO.int*. 15 Mar 2011. North Atlantic Treaty Organization. 02 Jan 2012 <http://www.nato.int/cps/en/natolive/topics_59378.htm>; “The North Atlantic Treaty,” *NATO.int*. 09 Dec 2008. North Atlantic Treaty Organization. 02 Jan 2012 <http://www.nato.int/cps/en/SID-A5F26964-970ECDE9/natolive/official_texts_17120.htm?selectedLocale=en>.
- 67 Cimbala and Forster 149.
- 68 Cimbala and Forster 97.
- 69 See Cimbala and Forster 97-148.
- 70 “Gen. Petraeus’ COIN Guidance,” *Afghanistan International Security Assistance Force*. 1 Aug. 2010. North Atlantic Treaty Organization. 12 Jan. 2011 <<http://www.isaf.nato.int/article/caat-anaysis-news/comisaf-coin-guidance.html>>.
- 71 Hendrickson 120.
- 72 Cimbala and Forster 167.
- 73 Cimbala and Forster 167; “History,” *Afghanistan International Security Assistance*

- Force*. North Atlantic Treaty Organization. 13 Jan. 2011 <<http://www.isaf.nato.int/history.html>>.
- 74 “Casualties: Afghanistan,” *CNN.com*. 13 Jan. 2011. Cable News Network. 13 Jan. 2011 <<http://www.cnn.com/SPECIALS/war.casualties/index.html>>; Amy Belasco, “The Cost of Iraq, Afghanistan, and Other Global War on Terror Operations Since 9/11,” *Congressional Research Service*. 29 March 2011. 21 Nov. 2011 <<http://www.fas.org/spp/crs/natsec/RL33110.pdf>>: 5.
- 75 Russell Goldman, “Obama Calls Afghan War ‘Difficult Endeavor,’ Withdrawal on Track for 2011,” *ABC News*. 16 Dec. 2010. 6 Jan. 2011 <<http://abcnews.go.com/Politics/president-obama-considerable-gains-made-afghan-warassessment/story?id=12412159>>.
- 76 Eric Schmitt, “Taliban Fighters Appear Blunted in Afghanistan,” *The New York Times*. 26 Dec. 2010. 24 Jan. 2011 <<http://www.nytimes.com/2010/12/27/world/asia/27policy.html>>.
- 77 “Remarks by the President on Osama Bin Laden,” *Office of the Press Secretary*. 2 May 2011. The White House. 2 May 2011 <<http://www.whitehouse.gov/the-press-office/2011/05/02/remarks-president-osama-bin-laden>>.
- 78 Elisabeth Bumiller, “Intelligence Reports Offer Dim View of Afghan War,” *The New York Times*. 14 Dec. 2010. 6 Jan. 2011 <<http://www.nytimes.com/2010/12/15/world/asia/15policy.html>>.
- 79 See Schmitt and “Gen. Petraeus’ COIN Guidance.”
- 80 Schweich; Carlotta Gall, “Afghanistan’s Opium Harvest Shrinks After Record Crop, U.N. Says,” *The New York Times*. 26 Aug. 2008. 6 Jan. 2011 <<http://www.nytimes.com/2008/08/27/world/asia/27afghan.html>>.
- 81 Richard A. Opiel, Jr., “Mysterious Blight Destroys Afghan Poppy Harvest,” *The New York Times*. 13 May 2010. 6 Jan. 2011 <<http://www.nytimes.com/2010/05/13/world/asia/13opium.html>>; Kirk Semple and Tim Golden, “Afghans Pressed by U.S. on Plan to Spray Poppies,” *The New York Times*. 8 Oct. 2007. 6 Jan. 2011 <<http://www.nytimes.com/2007/10/08/world/asia/08spray.html>>.
- 82 James G. March and Johan P. Olsen, *Rediscovering Institutions: The Organizational Basis of Politics* (New York: The Free Press, 1989) 55.
- 83 Craig Whitlock, “U.S. Troop Rollbacks May Not Stave off Calls for More Pentagon cuts,” *The Washington Post*. 9 Jan. 2011. 10 Jan. 2011 <<http://www.washingtonpost.com/wp-dyn/content/article/2011/01/09/AR2011010903630.html?sub=AR>>.
- 84 Judy Dempsey, “The Peril That NATO Can’t Ignore,” *The New York Times*. 10 Nov. 2010. 20 Jan. 2011 <<http://www.nytimes.com/2010/11/11/world/europe/11iht-letter.html>>.
- 85 “NATO and Libya – Operation Unified Protector.” *NATO.int*. 28 Mar 2011. North Atlantic Treaty Organization. 8 Apr. 2011 <http://www.nato.int/cps/en/natolive/topics_71652.htm?>.
- 86 “Operation UNIFIED PROTECTOR Final Mission Stats,” *NATO.int*. 2 Nov. 2011. North Atlantic Treaty Organization. 21 Nov. 2011 <http://www.nato.int/nato_static/assets/pdf/pdf_2011_11/20111108_111107-factsheet_up_factsfigures_en.pdf>.
- 87 “Turkey Backs NATO Command of Libyan operations,” *Al Jazeera*. 24 Mar 2011. 23 Nov. 2011 <<http://www.aljazeera.com/news/africa/2011/03/2011324185612655172.html>>; “France: NATO Will Support Libya Action After U.S.,” *Reuters*. 22 Mar. 2011. 8 Apr. 2011 <<http://www.reuters.com/article/2011/03/22/us-libya-france-idUSTRE72L2QA20110322>>; “NATO and Libya – Operation Unified Protector.”
- 88 “NATO and Libya: Operational Media Update,” *NATO.int*. 23 Oct 2011. North Atlantic Treaty Organization. 21 Nov. 2011 <http://www.nato.int/nato_static/assets/pdf/pdf_2011_10/20111024_111024-oup-update.pdf>.
- 89 Rick Gladstone, “U.N. Votes to End Foreign Intervention in Libya,” *The New York*

- Times*. 27 Oct 2011. 21 Nov. 2011 <http://www.nytimes.com/2011/10/28/world/middleeast/security-council-ends-libya-intervention-mandate.html?_r=1>.
- 90 Caldwell and Williams 43.
- 91 Caldwell and Williams 43.
- 92 “Treaty Between the United States of America and the Russian Federation on Further Reduction and Limitation of Strategic Offensive Arms (START I),” *Inventory of International Nonproliferation Organizations and Regimes Center for Nonproliferation Studies*. 27 July 2009. 23 Nov. 2011 <http://www.nti.org/e_research/profiles/Russia/Nuclear/treaties/start_1.html>.
- 93 “Strategic Concept.”
- 94 Peter Baker, “Senate Passes Arms Control Treaty With Russia, 71-26,” *The New York Times*. 22 Dec. 2010. 13 Jan. 2011 <<http://www.nytimes.com/2010/12/23/world/europe/23treaty.html>>; Karen DeYoung, “Ex-Eastern Bloc Nations Tell U.S. Senate to Ratify New START,” *The Washington Post*. 22 Dec. 2010. 13 Jan. 2011 <<http://www.washingtonpost.com/wp-dyn/content/article/2010/11/20/AR2010112002165.html>>.
- 95 John Steinbruner, “Security Policy and the Question of Fundamental Change,” *CISSM*. Nov. 2010. University of Maryland. 9 May 2011 <http://www.cissm.umd.edu/papers/files/security_policy_and_the_question_of_fundamental_change__nov_2010a.pdf>: 37-42; Vladimir Isanchenkov, “U.S.-Russian ‘Reset’ in Trouble as Nuke Pact Stalls,” *The Associated Press*. 19 Nov. 2010. 13 Jan. 2011 <<http://www.washingtontimes.com/news/2010/nov/19/us-russian-reset-trouble-uke-pact-stalls/?page=all>>.
- 96 “Strategic Concept.”
- 97 “NATO’s Nuclear Forces in the New Security Environment.” *NATO.int*. No date. North Atlantic Treaty Organization. 23 Nov. 2011 <http://www.nato.int/nato_static/assets/pdf/pdf_topics/20091022_Nuclear_Forces_in_the_New_Security_Environment-eng.pdf>.
- 98 “Strategic Concept.”
- 99 “NATO’s Nuclear Forces in the New Security Environment”; “Strategic Concept.”
- 100 Caldwell and Williams 44.
- 101 Caldwell and Williams 44; “Moscow Treaty Text,” *Department of Defense. Office of the Under Secretary of Defense for Acquisition, Technology and Logistics*. 9 Apr. 2011 <<http://www.acq.osd.mil/tc/treaties/sort/text.htm#1>>.
- 102 “WMD411: Extended Deterrence,” *James Martin Center for Nonproliferation Studies, NTI.org*. May 2010. North Atlantic Treaty Organization. 10 Jan. 2011 <http://www.nti.org/f_wmd411/f1b6_7.html>.
- 103 Isanchenkov.
- 104 Isanchenkov.
- 105 Baker, “Senate.”
- 106 Baker, “Senate.”
- 107 Baker, “Senate.”
- 108 Gordon Corera, “Loose Nukes Fear Spurs U.S.-Russian Action,” *BBC News*. 5 Jan. 2005. 6 Dec. 2010 <<http://news.bbc.co.uk/2/hi/europe/4149153.stm>>.
- 109 Peter Baker and Walter Pincus, “U.S.-Russia Pact Aimed At Nuclear Terrorism: Bush, Putin to Announce Plan to Counter Threat,” *The Washington Post*. 24 Feb. 2005. 6 Dec. 2010 <<http://www.washingtonpost.com/wp-dyn/articles/A48465-2005Feb23.html>>; For additional background into challenges with regard to reductions, see Steven Andreasen, Malcolm Chalmers, and Isabelle Williams, “NATO and Nuclear Weapons: Is a New Consensus Possible?” *Royal United Services Institute Occasional Paper*. Aug. 2010. RUSI. 23 Nov. 2011 <http://www.rusi.org/downloads/assets/NATO_and_Nuclear_Weapons.pdf>.

- 110 Backgrounder: Loose Nukes,” *Council on Foreign Relations*. Jan. 2006. 22 Nov. 2010 <<http://www.cfr.org/weapons-of-terrorism/loose-nukes/p9549>>.
- 111 J. Gregory L. Schulte, “NATO’s Nuclear Forces in a Changing World,” *NATO Review Web Edition 41.1* (1993). Last updated 22 Apr. 2002. North Atlantic Treaty Organization. 6 Dec. 2010 <<http://www.nato.int/docu/review/1993/9301-4.htm>>.
- 112 Ben Rhode, “WMD Proliferation,” *Europe and Global Security*, ed. Bastian Giegerich (London: Routledge, 2010) 156.
- 113 “Strategic Concept”; “Mission Statement,” IAEA. 23 Nov. 2010 <<http://www.iaea.org/About/mission.html>>.
- 114 See the preface of “Strategic Concept 2010” and Hans Blix, “A Season for Disarmament,” *The New York Times*. 4 Apr. 2010. 24 Mar. 2011 <<http://www.nytimes.com/2010/04/05/opinion/05iht-edblix.html>>.
- 115 David E. Sanger and William J. Broad, “U.S. Secretly Aids Pakistan Guarding Nuclear Arms,” *The New York Times*. 18 Nov. 2007. 24 Mar. 2011 <<http://www.nytimes.com/2007/11/18/washington/18nuke.html>>.
- 116 Rhode 156.
- 117 Rhode 163.
- 118 Rhode 175.
- 119 Rhode 166–167.
- 120 Robert F. Worth, “Report Undercuts Iran’s Defense on Nuclear Effort,” *The New York Times*. 8 Nov. 2011. 23 Nov. 2011 <<http://www.nytimes.com/2011/11/09/world/middleeast/irans-defense-of-nuclear-program-may-be-complicated-by-report.html>>.
- 121 “Strategic Concept.”
- 122 See Stephen Walt’s “Alliance Formation and the Balance of World Power.”
- 123 Amitav Acharya, “Collective Identity and Conflict Management in Southeast Asia,” *Security Communities*, eds Emmanuel Adler and Michael Barnett (United Kingdom: Cambridge University Press, 1998) 206. In-quote quotation from Alexander Wendt, “Collective Identity Formation and the International State,” *American Political Science Review* 88.2 (June 1994): 398.
- 124 Emmanuel Adler, “The Spread of Security Communities: Communities of Practice, Self Restraint, and NATO’s Post-Cold War Transformation,” *European Journal of International Relations* 14.2 (2008): 198.
- 125 Emmanuel Adler and Michael Barnett “A Framework for the Study of Security Communities,” *Security Communities*, eds Emmanuel Adler and Michael Barnett (United Kingdom: Cambridge University Press, 1998): 30.
- 126 Adler, “The Spread” 197.
- 127 “The North Atlantic Treaty,” *NATO.int*. 9 Dec. 2008. North Atlantic Treaty Organization. 28 Mar. 2011 <http://www.nato.int/cps/en/natolive/official_texts_17120.htm>; See Articles 1 and 2.
- 128 “The Consultation Process,” *NATO.int*. 8 Nov. 2010. North Atlantic Treaty Organization. 28 Mar. 2011 <http://www.nato.int/cps/en/natolive/topics_49187.htm>; Mary Hampton and James Sperling, “Positive/Negative Identity in the Euro-Atlantic Communities: Germany’s Past, Europe’s Future?” *Journal of European Integration* 24.4 (2002): 283. This article argues that transparency is a part of the relationship. The position taken here, however, is that self-restraint is the basis upon which other issues can be addressed, including transparency. Reciprocity on the basis of self-restraint is more likely to occur before transparency. One must occur before the other.
- 129 Adler and Barnett 31.
- 130 Adler and Barnett 54; “Istanbul Summit Communiqué,” *Istanbul*. 28 June 2004. Last updated 28 Nov. 2007. North Atlantic Treaty Organization. 16 Mar. 2011 <<http://www.nato.int/docu/pr/2004/p04-096e.htm>>.

- 131 Nigel Inkster, "Terrorism in Europe," *Europe and Global Security*, ed. Bastian Giegerich (London: Routledge, 2010) 99.
- 132 Veronica M Kitchen, "Argument and Identity Change in the Atlantic Security Community," *Security Dialogue* 40.1 (Feb. 2009): 106; Mällksoo 298.
- 133 Kitchen 97.
- 134 Kitchen 97.
- 135 Mällksoo 287.
- 136 Edward Said, *Orientalism* (London: Penguin Books, 2003) 7, 54.
- 137 Mällksoo 287.
- 138 Kitchen 106.
- 139 Kitchen 99, 106.
- 140 Kitchen 97; Cimbala and Forster 103.
- 141 See Kitchen 104–105.
- 142 See "NATO enlargement."
- 143 Adler, "The Spread" 200.
- 144 Soner Cagaptay, "Is Turkey Leaving the West?" *Foreign Affairs*. 26 Oct. 2009. 23 Mar. 2011 <<http://www.foreignaffairs.com/articles/65661/soner-cagaptay/is-turkey-leaving-the-west?page=show>>.
- 145 Hirofumi Shimizu and Todd Sandler, "Peacekeeping and Burden-Sharing, 1994–2000." *Journal of Peace Research* 39.6 (2002): 654–655; Jyoti Khanna, Todd Sandler, and Hirofumi Shimizu, "Sharing the Financial Burden for U.N. and NATO Peacekeeping, 1976–1996," *Journal of Conflict Research* 42.2 (Apr. 1998): 184. In each of these, this formula is used for peacekeeping.
- 146 Keith Hartley and Todd Sandler, "NATO Burden-Sharing: Past and Future," *Journal of Peace Research* 36.6 (1999): 665, 674.
- 147 Hartley and Sandler 670.
- 148 Hartley and Sandler 667.
- 149 For the most part, member states of Eastern Europe have demonstrated serious concerns about bellicose Russian behavior, especially after its invasion of Georgia in 2008. See Scott Shane, "NATO Balanced Baltic and Russian Anxieties," *The New York Times*. 6 Dec. 2010. 8 Apr. 2011 <<http://www.nytimes.com/2010/12/07/world/europe/07wikileaks-nato.html>>; "Financial and Economic Data Relating to NATO Defence," *NATO.int*. June 2010. 16 Nov. 2011. Table 3 "Defence expenditures as a percentage of gross domestic product." <http://www.nato.int/nato_static/assets/pdf/pdf_2010_06/20100610_PR_CP_2010_078.pdf>.
- 150 Daniel Fata, "Euro Defense Spending and NATO," *German Marshall Fund of the United States*. 20 Oct. 2009. 15 Mar. 2011 <<http://209.200.80.89/publications/article.cfm?id=678>>; "Financial and Economic Data Relating to NATO Defense."
- 151 Explicitly stated in "Strategic Concept" under "Defense and Deterrence."
- 152 See Hartley and Sandler 667.
- 153 "Troop Contributing Nations," *International Security Assistance Force*. 4 Mar. 2011. North Atlantic Treaty Organization. 16 Nov. 2011 <<http://www.isaf.nato.int/images/stories/File/Placemats/PLACEMAT.MARCH%2004..pdf>>.
- 154 Mark A. Boyer, "Trading Public Goods in the Western Alliance System," *The Journal of Conflict Resolution* 33.4 (Dec. 1989): 703.
- 155 Thom Shanker, "Obama Sends Armed Drones to Help NATO in Libya War," *The New York Times*. 21 Apr. 2011. 4 May 2011 <<http://www.nytimes.com/2011/04/22/world/africa/22military.html>>.
- 156 Anders Fogh Rasmussen, "NATO After Libya," *Foreign Affairs* 90.4. (2011): 2–6. Rpt. by *NATO.int*. 29 June 2011. North Atlantic Treaty Organization. 21 Nov. 2011 <http://www.nato.int/cps/en/natolive/opinions_75836.htm?selectedLocale=en>.
- 157 Kitchen 105; Cimbala and Forster 166.

- 158 Mark Webber, "NATO: The United States, Transformation and the War in Afghanistan," *The British Journal of Politics and International Relations* 11.1 (2009): 50.
- 159 Henry Chu, "Britain Plans Military Cutbacks," *Los Angeles Times*. 20 Oct. 2010. 21 Nov. 2011 <<http://articles.latimes.com/2010/oct/20/world/la-fg-britain-military-20101020>>; "Troop Contributing Nations."
- 160 Chu.
- 161 See Thom Shanker and Christopher Drew, "Gates Sees Crisis in Current Spending," *The New York Times*. 14 Feb. 2011. 16 Mar. 2011 <<http://www.nytimes.com/2011/02/15/us/politics/15pentagon.html>>; Robert Pear, "Budget Stalemate Leaves Chaos at Many Agencies," *The New York Times*. 14 Mar. 2011. 16 Mar. 2011 <<http://www.nytimes.com/2011/03/15/us/15spend.html>>.
- 162 Indeed, it was evident even during the 1970s that economics was a major concern for states, as well. The Oil Crisis is a primary example of this; Joseph S. Nye, Jr. and David A. Welch, *Understanding Global Conflict and Cooperation: An Introduction to Theory and History* (New York: Longman, 2009) 259-260.
- 163 Rasmussen.
- 164 Alyson K. Bailes, "NATO's European Pillar: The European Security and Defense Identity," *Defense Analysis* 15.3 (1999): 305.
- 165 Bailes 307-308.
- 166 Laurent Pirot, "France Officially Asks to Rejoin NATO command," *USA Today*. 20 Mar. 2009. 17 Mar. 2011 <http://www.usatoday.com/news/world/europe/2009-03-20-2217142846_x.htm>; Neil Winn, "Towards a Common European Security and Defense Policy? The Debate on NATO, the European Army and Transatlantic Security," *Geopolitics* 8.2 (2003): 56; Hampton and Sperling 285.
- 167 Bailes 311, 309; Hampton and Sperling 289.
- 168 Bastian Giegerich, "Military and Civilian Capabilities for EU-led Crisis Management Operations," *Europe and Global Security*, ed. Bastian Giegerich (London: Routledge, 2010) 44-45.
- 169 Giegerich 41, 47.
- 170 Steven Erlanger, "Europe's Foreign Policy Chief, Struggling for Mandate, Faces Criticism on Uprising," *The New York Times*. 10 Feb. 2011. 19 Apr. 2011 <<http://www.nytimes.com/2011/02/11/world/europe/11europe.html>>.
- 171 See Wallander 706-707.
- 172 In the heat of an American election cycle, the fate of the defense budget remains to be seen.
- 173 Rasmussen.
- 174 "Senate Committee on Appropriations on the Fiscal 2012 Department of Defense." *Joint Chiefs of Staff*. 15 June 2011. 22 Nov. 2011 <<http://www.jcs.mil/speech.aspx?id=1617>>.
- 175 Rasmussen.

BIOGRAPHIES

Editor

melanie dell

MELANIE DELL, RECIPIENT of the UMBC Alumni Association Legacy Scholarship and President's Fellows Award, is graduating in May 2012 with a B.A. in English and a minor in French. Serving as a *UMBC Review* editor for the past two years has been a rewarding and enlightening experience for her. She plans to apply the skills she has learned from this job when she pursues a Master of Library and Information Science degree in Fall 2012. Melanie would like to thank her co-editor, Esther Gross, for her excellent editing abilities and enthusiasm for the job. She is also forever appreciative of the *UMBC Review's* faculty advisor, Dr. Falco, who has provided her with not only much help in this job but also endless support and guidance concerning her academic future. Melanie would also like to express thanks to the Office of Undergraduate Education, especially Janet McGlynn. Melanie is sad that her time is ending with the *UMBC Review*, but she looks forward to the new academic experiences that await her.

Editor

esther gross

ESTHER GROSS IS a MARC U*STAR Scholar in her third year at UMBC. This is her first year working as *UMBC Review* editor. As a mathematics major, she has really enjoyed the increased exposure to her own discipline as well as that of others. Her goal is to pursue a PhD in mathematics and eventually become a professor. The *UMBC Review* has taught her not only about editing but also about the teamwork required to publish a research journal. She would therefore like to thank the senior editor, Melanie Dell, for the countless hours that she worked and her never-ending willingness to help. Esther would also like to thank Dr. Falco for all of his encouragement and advice that was often accompanied by cookies. She extends her thanks to the Office of Undergraduate Education. Without their funding, this journal would not be possible. Lastly, she would like to thank Janet McGlynn for her attention to detail and encouragement throughout the review process. Esther has been grateful for this opportunity and looks forward to next year's edition.

Designer

brendan lipton

BRENDAN LIPTON, RECIPIENT of the President's Fellows Award and member of the Honors College, is graduating in May 2013 with dual degrees, a BA in Graphic Design and a BA in Media and Communications Studies. Working with the *UMBC Review* editors has been an immensely rewarding experience and he appreciated the knowledge he has gained on the complexities of book design from Professor Guenet Abraham. He is thankful to be able to join the ranks of designers who have been privileged to design the *UMBC Review*. Brendan would also like to express his gratitude to his co-designer Faryal Khalid for all of her assistance in making the design process smooth and enjoyable. He also extends his thanks to the Office of Undergraduate Education without whose support the publishing of the *UMBC Review* would not be possible. It is his sincere hope that you enjoy reading the 2012 edition of the *UMBC Review*.

Designer

faryal khalid

FARYAL KHALID, RECIPIENT of the Dean's Scholarship Award and a member of the Omicron Delta Kappa Honors Fraternity, is graduating in May 2012 with a BA in Graphic Design. Getting the opportunity to design the *UMBC Review* has been momentous in her career as an undergraduate student at UMBC. Being mentored by Guenet Abraham, the *UMBC Review* Faculty Advisor for design, has been one of the most meaningful learning experiences. Faryal is very grateful for the constant support of her co-designer, Brendan Lipton. His professional attitude and strong work ethic has made this partnership very beneficial and successful. She would also like to extend her gratitude to the Office of Undergraduate Education, Melanie Dell and Esther Gross for their continuous support. She hopes that the hard work of so many people results in a journal that is enjoyed by all.

