SURJ

Volume 4
Spring 2005

STANFORD UNDERGRADAUTE
RESEARCH JOURNAL

The 2004-2005 issue of the *Stanford Undergraduate Research Journal* (SURJ) represents a tremendous turning point for SURJ, both as a student group and as a publication. Our fourth year in existence has been one of incredible growth and new challenges, and our continued success is a testament to the strength of the Stanford community in all its varied forms.

Building off last year's successes, SURJ has continued to promote open access to research and ideas. We continue to widely distribute journals both across campus and throughout the nation, and have expanded our online presence through cross-linkage with journals at peer institutions. This year has been one of fostering relationships—both within the Stanford community and with the outside world. On campus, SURJ has striven to increase its collaboration with departments and the administration. As a result of our cooperation, the Office of Undergraduate Admissions has begun sending issues to newly admitted students. Across the nation, SURJ serves to excite future Stanford about the possibilities of discovery during their undergraduate years. Our close relationship with the Undergraduate Research Program (URP) office has allowed us to promote SURJ across campus to students, parents, and faculty.

Outside of Stanford, SURJ has strengthened its relationships with journals at peer institutions as we attempt to promote the sharing of knowledge. SURJ has played the role of both mentor for and collaborator with many of these journals as they attempt to build their own publications. We hope to keep developing these relationships in the future. The journal's reputation has made it as far as Europe. This year SURJ's staff was visited by students from the Radboud University Nijmegen in the Netherlands, who are starting a journal of their own. This type of international information sharing only serves to inspire SURJ and strengthen our confidence in the importance of sharing student research.

On behalf of the 2004-2005 SURJ staff and authors, we'd like to thank you for taking the time to read this publication. SURJ attempts to showcase undergraduate research at its best, in all its varied forms, and we hope you find the issue as edifying and entertaining as we do. Without the high quality of research conducted by Stanford undergraduates, the hard work of our staff, and the continued support of both faculty and administration, SURJ would not have nearly the reputation and campus presence that it enjoys today. SURJ has slowly grown into a mature publication, and we hope that your continued support will allow us to thrive in the future.

Sincerely,

Kevin K. Gao and Peter M. Pawlowski
Editors-in-Chief, 2004-2005

SURJ Staff

Biology

Design, Synthesis, and Evaluation of Dihydroisoxazole Analogs as Irreversible Inhibitors of Tissue Transglutaminase

Eun Cho

Coeliac Sprue disease is caused by presence of gluten peptides from common food grains like wheat and barley in genetically susceptible individuals. Unfortunately, therapeutic reagents have not been developed other than strict, lifelong gluten-free diet. Recent studies have shown that enzyme tissue transglutaminase (TG2) performs a crucial role in the disease mechanism by recognizing and modifying the gluten peptides, which is necessary for the onset of a complex inflammatory response. By closely mimicking these proteolytically stable gluten peptides, successful analogs that bind to selectively inhibit TG2 from interacting with the gluten peptides have been found. From structural modification of these analogs leaving the reactive moiety dihydroisoxazole intact, we have synthesized and tested in vitro the lead compound from this series. Incorporation of dioxobenzothiophene group and 5-hydroxytrptophan to dihydroisoxazole significantly increased the inhibition of TG2. Further studies on the stabilization and the mechanisms of this compound should lead to discovery of more potent inhibitors against tissue transglutaminase and possibly, therapeutic agents against Coeliac Sprue.

At 76 kD with 686 amino acids, tissue transglutaminase (TG2) is a calcium-dependent multifunctional enzyme that catalyzes post-translational modification of specific glutamine residues. TG2 is a ubiquitous enzyme functioning at various cell locations and plays important roles in many cellular processes including cell death, cell movement, adhesion, and proliferation. Even though TG2 possesses many protective functions in humans, its defective expression can lead to diverse clinical disorders. Neurodegenerative disorders, such as Huntington's, Alzheimer's, and Parkinson's diseases, are linked with TG2-catalyzed transamidation reac-tions whereas diseases like Coeliac Sprue are related to the deamidation activity of the enzyme.

Despite its importance in mammalian biology, animal study has shown that ablation of both copies of the TG2 gene does not result in an embryonic lethal phenotype. In fact, TG2-/TG2- animals are viable with normal size and weight and are born with Mendelian frequency[1]. A possible explanation for this paradox is that other types of transglutaminases might compensate for the absence of TG2 in mammalian tissues[2]. Although the knockout animals show some type of impaired wound healing, autoimmunity, and diabetes, the relatively low toxic effect of TG2 deletion is clear from this study. In addition, it can be inferred that TG2 inhibitors are potential pharmacological agents to treat some of the diseases described above.

Coeliac Sprue is of particular interest; even though coeliac disease is prevalent among Caucasians, affecting 1 in every 200 individuals, no effective treatments have been developed other than strict, lifelong gluten-free diet. This malabsorption syndrome caused by intolerance to ingested gluten proteins from common food grains like wheat, rye, and barley is characterized by total atrophy of villi and flattening of the mucosa from chronic inflamma-

tion in the jejunum of small intestine. Some of the clinical symptoms include diarrhea, malabsorption of nutrients, weight loss, abdominal distension, anemia, osteoporosis, and intestinal malignancies[3].

The disease mechanisms of complex inflammatory disorders depend on extensive interactions between genetic and environmental factors. In coeliac disease, gluten is the environmental factor, and HLA molecules are the genetic factors. The strong indication of coeliac disease as a genetic disorder comes from the high sibling relative risk of the disease, ranging from 20 to 60[4]. Furthermore, genetic studies of the patients show that HLA genes are involved in coeliac disease. These genes transcribe and translate into antigen-binding receptors on the surface of antigen-presenting cells (APC) in the small intestine. Coeliac disease patients express a particular HLA-DQ2 heterodimer (DQ8 for minority); they have either DR3-DQ2 haplotype, or are DR5-DQ7 and DR7-DQ2 heterozygous[4]. The DQ2 molecules bind and present the gluten peptides to T cells, causing inflammation in the small intestine.
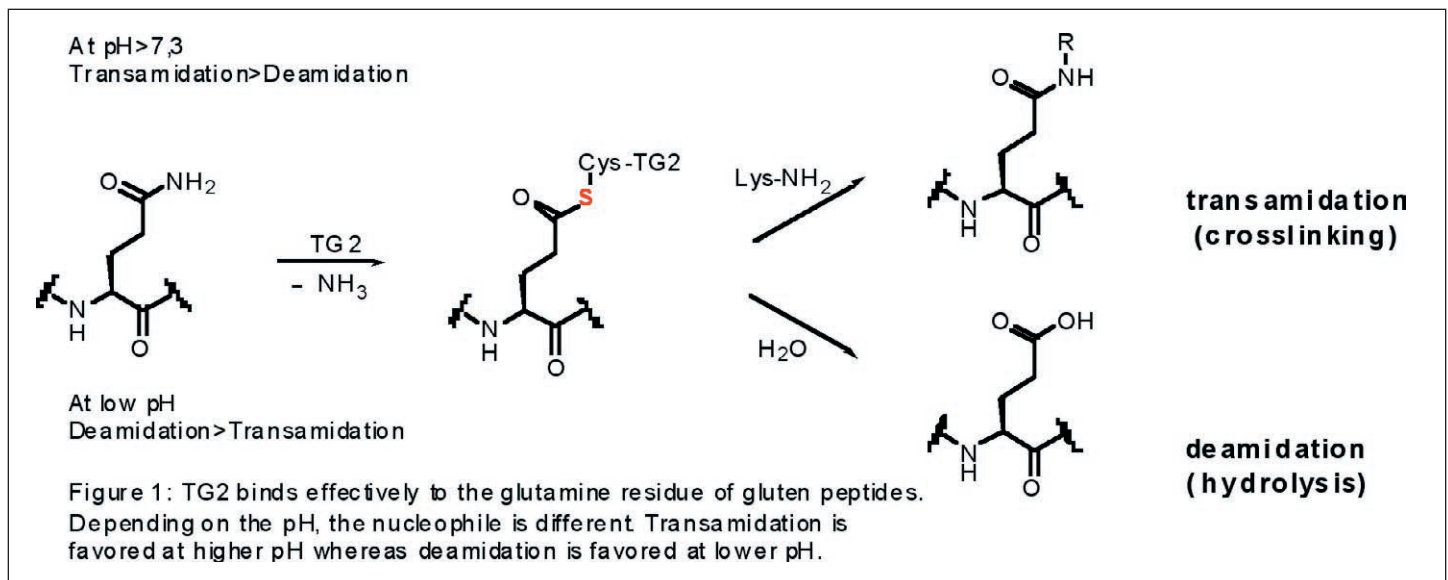
The binding of gluten peptides to HLA molecules rely on several factors. As described above, coeliac disease patients have particular DQ2 molecules that can bind efficiently to gluten peptides. Gluten peptides are resistant to processing by gastric, pancreatic, and intestinal brush-border enzymes because of the high content of glutamine and proline amino acids. It is postulated that instead of being degraded into individual amino acid, these gluten peptides are transported across the mucosal epithelium as polypeptides. Another important factor involves the activity of tissue transglutaminase. The deamidation by TG2, which is favored over the transamidation in the proximal small intestine at pH around 6.6 (see Figure 1), allows gluten peptides to bind tightly to HLA molecules that have preference for negatively charged residues[4]. In addition, gluten peptide-HLA molecule complex on the surface of antigen-presenting cells can in turn, interact more efficiently with CD4+ T cell receptors in the lamina propia of the small intestine, leading to a complex inflammatory response and villous atrophy.

Since the pathway of coeliac disease depends on numerous factors, various types of therapeutic methods have been proposed to treat the disease.

As one of these approaches, this paper presents selective inhibition of tissue transglutaminase to prevent deamidation of gluten peptides. Without deamidation, the interaction between gluten peptides and DQ2 or DQ8 molecules on APCs would not occur efficiently. Therefore, CD4+ T cells would not be stimulated to trigger the complex inflammatory response necessary for the onset of Coeliac Sprue.

Many inhibitors of TG2 have been designed and synthesized in our laboratory. Observation and evaluation of gluten peptides' resistance to proteolysis as well as their high affinity to TG2 have led to the development of successful inhibitors, such as acivicin and 3-halo-4,5-dihydroisoxazole. The basic design principle of these inhibitors is to mimic the glutamine residue of TG2 substrate peptides. Specifically, these inhibitors inactivate tissue transglutaminase by covalent modification of its reactive group; the cysteine thiol group of TG2 forms a covalent bond with the dihydroisoxazole by replacing the halogen atom (see Figure 2). After an inhibitor molecule binds to the enzyme and blocks the enzyme turnover, gluten peptides would not associate with TG2 anymore. Therefore, the gluten peptides would not be deamidated

Figure 1: TG2 binds effectively to the glutamine residue of gluten peptides. Depending on the pH, the nucleophile is different. Transamidation is favored at higher pH whereas deamidation is favored at lower pH.

to interact with the HLA molecules on the antigen presenting cells.

As part of the on-going efforts to develop dihydroisoxazole-based TG2 inhibitors, analogs of KCC009 (see Figure 3) have been synthesized in our laboratory and tested against human tissue transglutaminase for their inhibition activities. The kinetic assay results showed some important properties of KCC009 (unpublished data). First, bulky amino acid side chain of the inhibitor was important for the activity, such as in the tyrosine and tryptophan derivatives. In addition, the stereochemistry of the side chain was important, as only L-amino acid version of the inhibitor yielded high activity against TG2. The secondary amide linking dihydroisoxazole to tyrosine was necessary since changing it into either an oxygen atom or tertiary amide produced low activity. The study also showed importance of the bulky group on the N-terminal.

Using KCC009 as a benchmark inhibitor because of its moderate activity against tissue transglutaminase (kinh/KI=2000) and good water solubility (1.5 mM in 5% DMSO), we focused on modification of the carbamate group to ascertain a new lead compound in the series.

The aim of this paper is to summarize the design and synthesis of various dihydroisoxazole analogs, to discuss the kinetic assay results of the newly synthesized inhibitors, and to suggest follow-up studies that could lead to discovery of more potent inhibitors against tissue transglutaminase.

Materials and Methods

Tissue Transglutaminase Expression and Purification

The recombinant type of TG2 used in the experiment is bigger than the wild-type with its molecular weight of 79 kD with 707 amino acids. This is because His-tag addition is necessary for efficient purification of the enzyme. First, BL21 Escherichia coli inoculum was grown in 37^0C shaker for 18 hours. Then, 5 mL of the seed culture was transferred to 1 L of LB medium containing kanamycin. After inducing the culture with 1 μM IPTG, the cell culture was incubated in 18^0C shaker for 20 hours. Sonification was used for cell lysis. Elution was carried out through Ni-NTA column using 100 mM imidazole. Purity of each fraction was checked by electrophoresis. After exchanging buffer, glycerol was added to 10% of the final volume. The purified enzymes were aliquoted and flash-frozen in liquid N2 to be stored in -80^0C freezer until use.

General Scheme of Synthesis

The organic synthesis of dihydroisoxazole inhibitors comprised of three major steps, the first one being preparation of 4-nitrophenyl carbonates. This step involved a reaction between an alcohol and either 4-nitrophenyl chloroformate or bis-(4-nitrophenyl) carbonate depending on the type of alcohol used. By varying the alcohol group, different analogs of dihydroisoxazole inhibitors with different carbamate groups were able to be synthesized at the end. Most of the reagents were obtained from Sigma-Aldrich. 1,1-Dioxobenzo[b]thiophen-2-ylmethyl chloroformate was obtained from Fluka whereas 2-pyridyl carbinol, 2-naphthalene methanol, and 2-(phenyl-sulfonyl) ethanol were acquired from ASDI. The second part of the synthesis was the preparation of carbamate-protected tyrosines. Although the reaction between dioxobenzo[b]thiophen -2-ylmethyl chloroformate and tyrosine was successful, carbonates formed from the previous step were reacted with tyrosine esters for higher yield. Based on the circumstances, different tyrosine ester compounds were used to be worked up. The final step involved the introduction of dihydroisoxazole to the carbamate-protected tyrosines using EDCl and HOBt as coupling reagents. Between each major step, column chromatography was carried out for purification, and the purity and the structure of the product compound were checked by NMR spectroscopy from Varian Gemini-200, Gemini-300, or Mercury-400 with tetramethylsilane as a reference.

Figure 2: Dihydroisoxazole inhibitors' mode of inhibition against TG2 is the covalent bond formation between the ring and the reactive cysteine thiol group of the enzyme.
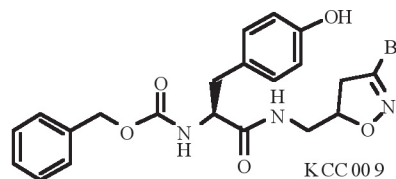

Figure 3: KCC009 is a dihydroisoxazole inhibitor with moderate activity and solubility in water. This compound was chosen as the benchmark inhibitor.
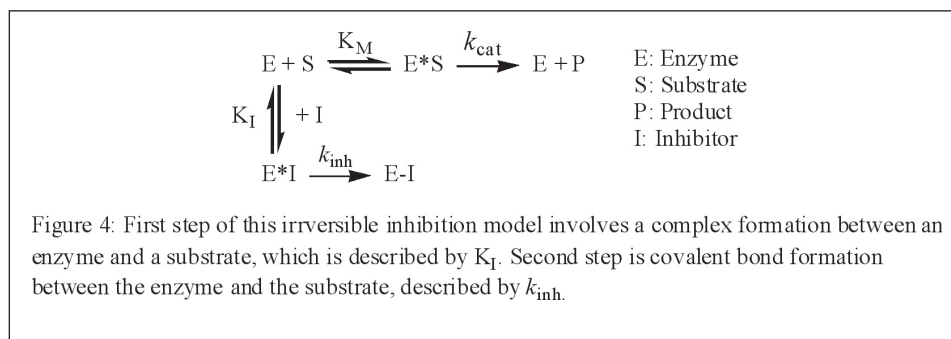
Kinetic Assay for Tissue Transglutaminase

Newly synthesized dihydroisoxazole compounds were assayed against recombinant human TG2 using an enzyme coupled assay, where glutamate dehydrogenase was used to couple ammonium ion, a product of the TG2-catalyzed reaction, to α-ketoglutarate to yield glutamate and the simultaneous oxidation of NADH to NAD+, which could be monitored at 340 nm by UV spectroscopy to follow the reaction progress. Enzymes for biochemical assays were obtained from Sigma-Aldrich.

Compounds were assayed in a reaction mixture containing 200 mM MOPS (pH=7.2), 5 mM CaCl2, 1 mM EDTA, 10 mM α-ketoglutarate, 18 U/mL glutamate dehydrogenase, 0.4 mM NADH, 3.3 % (v/v) DMSO, 0.5 μM TG2, and 2, 5, 10 , or 20 mM Cbz-Gln-Gly·Na (KM=11 mM, kcat=37 min-1). TG2 was introduced at the end, 3 minutes after glutamate dehydrogenase addition. The enzyme reaction was started, and the consumption of NADH was monitored by UV spectroscopy at 340 nm.

Results and Discussion

After kinetic assay of the dihydroisoxazole analogs with TG2, kinetic parameters were acquired by plotting the reaction progress curve against theoretical equations of the following irreversible inhibition model[3, 5]. With inhibitor concentration being much greater than enzyme concentration, specificity (kinh/KI) was related to the relative efficiency of each inhibitor.

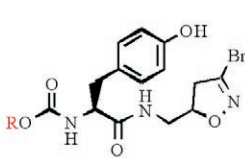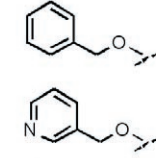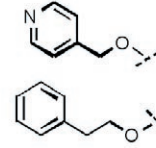$$E + S \xrightleftharpoons{K_M} E*S \xrightarrow{k_{cat}} E + P$$

$$K_I \updownarrow \, + I$$

$$E*I \xrightarrow{k_{inh}} E\text{-}I$$

E: Enzyme
S: Substrate
P: Product
I: Inhibitor

Figure 4: First step of this irrversible inhibition model involves a complex formation between an enzyme and a substrate, which is described by K_I. Second step is covalent bond formation between the enzyme and the substrate, described by k_{inh}.
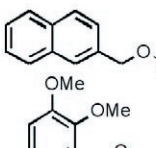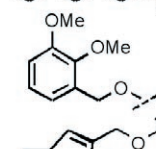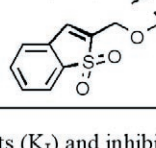
	K_I (mM)	k_{inh} (min^{-1})	k_{inh}/K_I (min^{-1}M^{-1})	
	0.42	0.86	2000	KCC009
	0.078	0.21	2700	E015
	0.081	0.12	1500	E056
	0.061	0.093	1500	E047
	0.043	0.070	1600	E040
	0.067	0.12	1800	E014
	0.087	0.38	4400	E009

Table 1: Dissociation constants (K_I) and inhibition rate constants (k_{inh}) of dihydroisoxazole inhibitors of human TG2 were obtained from irreversible inhibition model. Specifici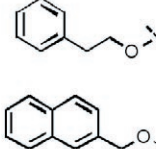ty (k_{inh}/K_I) was related to relative efficiency of each inhibitor. Introduction of nitrogen atom to the 3 position of the benzene ring (E015) and and replacement with dioxobenzothiophenyl group (E009) increased the inhibition activity against TG2 compared to that of KCC009.

Carbamate Group Variation of KCC009

Several analogs differing in the carbamate group were synthesized to assess the structure-activity relationships of KCC009. Introduction of nitrogen atom to the 3 position of the benzene ring increased the inhibition activity. The assay results showed that E015 is a potent inhibitor of human TG2, its specificity being 2700 min-1 M-1. However, introduction of nitrogen atom to the 4 position of the benzene ring decreased the inhibition activity relative to that of KCC009. Transition from the benzyl to phenyl ethyl group also diminished the inhibition activity as shown with E047. Changing the benzyl to bulkier naphthyl group lessened the specificity. However, when a bulkier dioxobenzothiophenyl group replaced the benzyl group of KCC009, the inhibition activity against TG2 was markedly improved, which is clear from Figure 5. It is unclear why such observations were made with the modification of KCC009, and further studies and research in the structure of the TG2 binding site will be needed. With its specificity around 4400 min-1 M-1, modification of E009 was investigated.

Variation of E009

The dioxobenzothiophenyl group was changed to the phenyl sulfonyl group to assess whether the sulfonyl group had to be placed in a ring for the efficient inhibition against TG2. E073 had even lower specificity than that of KCC009, with the specificity being close to 1500 min-1 M-1. E075, similar to E009 without the introduction of dihydroisoxazole, was synthesized in order to ascertain whether the dioxo-benzothiophene could be another reactive site to TG2. Enzyme inhibition with the compound was absent (see Figure 6); the dihydroisoxazole group was the main site that interacted with the cysteine thiol of TG2. Based on this result, the dioxobenzothiophene group was fixed, and the side chain of the amino acid moiety was varied using different amino acids. Tryptophan and 5-hydroxytryptophan were chosen since replacing tyrosine in KCC009 with either of the two amino acids produced higher inhibition activity against TG2 (unpublished data).

E081, tryptophan analog of E009 was first synthesized and assayed. However, v0 was higher than that without the inhibitor, and the kinetic parameters could not be obtained. The data implied that the compound was initially aiding substrate association with TG2. Its inhibition activity was not better than that of E009. Exchanging tyrosine with 5-hydroxy-tryptophan dramatically increased the specificity of the dihydroisoxazole inhibitor, from 4400 min-1 M-1 to 6800 min-1 M-1. This extent of increment in the specificity was not seen with KCC009 variation. A possible explanation for this phenomenon is the synergic effect on the orientation and proximity of the inhibitors to the reactive site of TG2; effect of dioxo-benzothiophen and 5-hydroxytrypto-
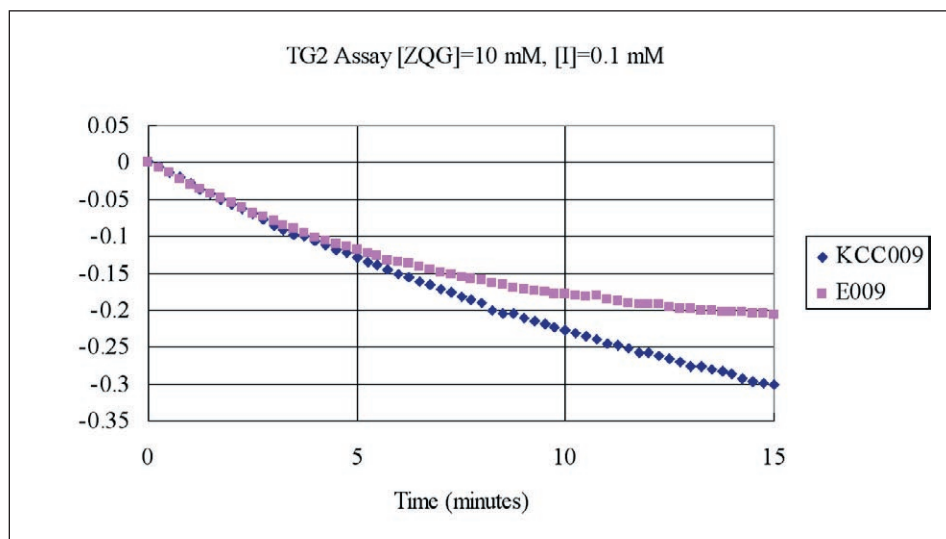
Figure 5: This graph from the enzyme coupled assay clearly showed that E009 had better inhibition activity against TG2 than KCC009 with its faster plateau. NADH had UV absorbance at 340 nm whereas NAD+ did not, and TG2 was responsible for the oxidation of NADH to NAD+ in the kinetic assay. Delta absorbance was computed by subtracting the initial UV absorbance value from the data. There was reduced delta absorbance at 340 nm when E009 was introduced compared to that of KCC 009. This was because NADH could not be oxidized to NAD+ with more powerful TG2 inhibition. Multiple trials of the kinetic assay indicated that the data was reproducible.
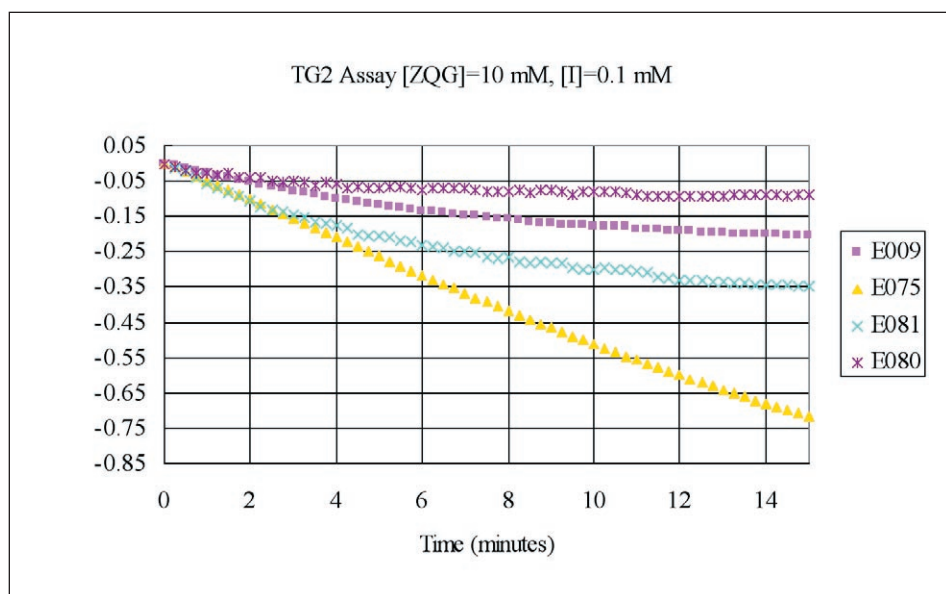
Figure 6: UV spectroscopy data at 340 nm showed that E080 has the highest inhibition activity against human tissue transglutaminase. With more powerful inhibition of TG2, NADH could not be reduced to NAD+, as indicated by the faster plateau in the graph. Inhibition of TG2 with E075 and E081 were absent. Multiple trials of the kinetic assay were done to test the reproducibility of the data.

phan presence together is greater than the sum of their individual effect on the inhibition activity against human tissue transglutaminase.

E080 has the highest specificity out of the dihydroisoxazole inhibitors up to date by a large margin. Water solubility of E009 and E080 is less than that of KCC009 (1.5 mM in 5% DMSO) by a factor of 2 (unpublished data). Lower solubility in water may be disadvantageous in the context of drug delivery.

Conclusions

With the highest specificity (kinh/KI) among the dihydroisoxazole inhibitors, compound E080, a dioxo-benzothiophenyl analog of KCC009, has been discovered as a potent inhibitor of human tissue transglutaminase. Additional studies on E080 should be done to check its stability, which is crucial for further studies using this compound. Dioxobenzothiophene might be a problematic site that could react with other types of nucleophiles. Structural modification of E080 may be needed in that case. Although drug delivery to small intestine is relatively easy, the inhibitor must survive degradation along pathway to maximize its uptake. Other possible experiments include cell-based assays to check the compound's mode of action inside cells and in vivo experiments to obtain information concerning bioavailability and potency of the inhibitor.

Besides TG2 inhibition, which was the focus of this paper, there are other potential targets to coeliac disease treatment. One way is to silence gluten-specific T cells to prevent its stimulation leading to a complex inflammatory response. Digestive destruction of T cell stimulating gluten peptides is another promising approach. Blocking of peptide presentation on HLA molecules is also feasible.

	K_I (mM)	k_{inh} (min^{-1})	k_{inh}/K_I (min^{-1}M^{-1})	
	0.087	0.38	4400	E009
	0.16	0.24	1500	E073
	--	--	--	E075
	--	--	--	E081
	0.079	0.54	6800	E080

Table 2: Dissociation constants (K_I) and inhibition rate constants (k_{inh}) of E009 analogs of human TG2 were obtained from irreversible inhibition model. Reduced specificity value with E073 showed that sulfonyl group had to be kept in a ring. Kinetic data from E075 showed that the dihydroisoxazole group was the main site that interacted with the cysteine thiol of TG2. There was a significant increase in the TG2 inhbition when tyrosine group was replaced with 5-hydroxytryptophan.

References

1. Laurenzi VD, Melino G.. Gene disruption of tissue transglutaminase. Mol Cell Biol 2001; 21: 148-155.

2. Fesus L, Piacentini M. Transglutaminase 2: an enigmatic enzyme with diverse functions. Trends Biochem Sci 2002; 27: 534-539.

3. Hausch F, Halttunen T, Maki M et al. Design, synthesis, and evaluation of gluten peptide analogs as selective inhibitors of human tissue transglutaminase. Chem Biol 2003; 10: 225-231.

4. Sollid LM. Coeliac disease: dissecting a complex inflammatory disorder. Nat Rev Immunol 2002; 2: 647-655.

5. Chittur SV, Klem TJ, Shafer CM et al. Mechanism for Acivicin Inactivation of Triad Glutamine Amidotransferases. Biochemistry 2002; 40: 876-887.

Acknowledgements

Eun Cho

Eun was born in Seoul, Korea and is currently a junior majoring in chemistry and biological sciences/cell biology. Eun was awarded the Stanford Chemistry Undergraduate Bing Research Fellowship which allowed him to research under the mentorship of Professor Khosla. Eun plans to pursue a career in the biotechnology industry. Eun would like to thank his family and friends for their continual support.

Computer Science

Publishing Content on the Web
Content Management fitting any structure

Isil Ozgener and Thomas Dillig

We propose a method for dynamic content management of large and structurally heterogeneous websites. The content management system (CMS) discussed in this paper allows even non-specialists to easily update webpages and add new ones without knowing anything about the details of the HTML language. The database driven implementation of this CMS adopts a recursive structure that constructs a complex web page using structural units like a layout table to arrange the elementary atomic units of information. This CMS completely separates content from structure and always guarantees well-formed websites, including a well-formed navigation structure. This system is currently used by the Carnegie Institution's Department of Plant Biology as a tool for expanding their project website and for incorporating new experimental data with highly heterogeneous structure.

Many organizations need to provide accurate and up-to-date information on their web pages. While this task is relatively straightforward for micro sites with fewer than 10 pages, large and frequently revised websites benefit greatly from the use of a content management system (CMS) [1]. A CMS allows people to easily enter or update information without knowing anything about the details of HTML or webpage implementation.

Most of the well-known tools ([2], [3] and [4]) for web content management target very large and structurally uniform sites such as on-line stores. In such websites, every product has certain pieces of information attached to it, such as a price tag, a picture, and a description, making the structure uniform and allowing a single pre-prepared template to fit each product [5]. Many commercial tools are available for updating such structurally uniform websites, but there are no convenient means of managing content when the nature of the information is variegated and when future revisions are unforeseeable.

Problems associated with a lack of CMS for structurally heterogeneous websites arise especially for scientists who need to share newly-acquired experimental data through their websites. For instance, biologists often need to conveniently post downloadable DNA sequences and new microarray data on their website to allow new experimental data to be shared among collaborators and fellow specialists. They also need to incorporate tools for comparative genomics into their websites. In cases like this where the information posted on the web has no homogeneous structure, expert webpage implementers must be consulted before any new information can be shared via the Internet. However, a CMS to handle structurally-heterogeneous content eliminates this need and allows research groups and business organizations to enter new information conveniently and without significant delay.

In this paper, we discuss the implementation of a CMS that works also for structurally heterogeneous websites. This CMS is based on atomic units of information (AUI) nested hierarchically. A webpage is constructed by nesting together many different kinds of AUI's, where each kind corresponds to a unique structural scheme. This technique of using hierarchically-arranged AUI's helps to differentiate between distinct kinds of structural information and allows content to be entered within a strictly hierarchical relationship that makes

it almost impossible even for a non-expert to delete crucial information from the website.

Content on One Page

In our implementation of this CMS, we differentiate between five distinct types of AUI that compose a typical webpage: Text, Image, Hyperlink, Downloadable File, and Scripting Module. Text represents plain HTML text with a specified format; Image allows the user to insert image files with an optional caption. Downloadable files are any binary or text data download such as a PDF file. Finally, Scripting Modules are any script (i.e. code) written to perform a special function such as displaying a news ticker.

Any information on one given page can be represented as a collection of these AUI's. However, in order to structure these elements we need to define relationships between them. The relationship between these AUI's is defined via the atomic unit of structure (AUS) within one page. An AUS, which we more informally name a "layout table", groups zero or more AUI's or AUS's together and defines the displayed layout of the information. This grouping, however, is only uni-level, resulting in at most one other object per cell. Graphically, each AUS corresponds to a graph node with one parent and multiple (possibly
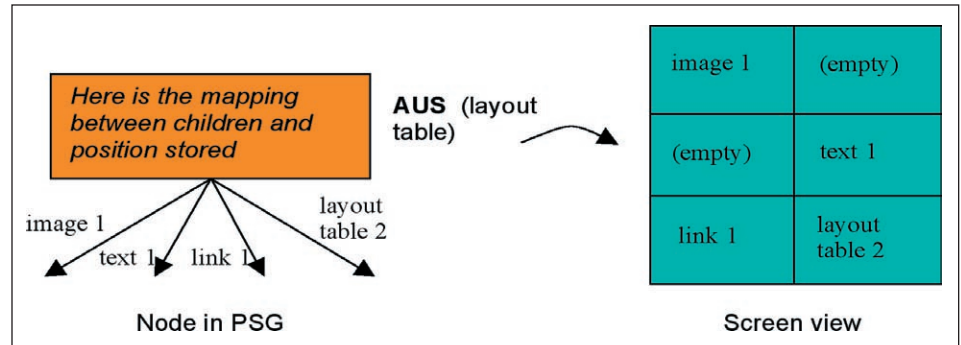
Figure 1: An AUS such as a layout table defines how information is displayed on the webpage, where each piece of information can be an AUI or another AUS.

zero) children, where all children are mapped to fixed screen layout. The exact nature of this mapping needs to be defined for each AUS. For instance, layout tables map each child to its row/column position on the screen (see Figure 1).

Therefore, AUS's and AUI's recursively define a Page Structure Graph (PSG), which represents the hierarchical content structure displayed on a webpage. The PSG for any well-formed structure is always guaranteed to be a tree. The root of this tree is a special node containing a list of objects – i.e AUI's or AUS's- , mapped to a simple consecutive screen layout (see Figure 2). Effectively, the root of the PSG is a default AUS that allows only one column and has as many rows as the number of its children. The leaves of a PSG are always AUI's. To add a new object to the PSG, it's sufficient to state its parent to uniquely determine

its place in the PSG. The default parent of any newly created object is the root node of the PSG. The implementation of the PSG as a tree structure makes it especially easy to move content within a page from the user's perspective. The user only needs to specify a new parent for the object to reorganize the PSG and hence the displayed layout of the webpage. As discussed later, this approach generalizes across page boundaries. Figure 2 gives an example of a typical PSG. Here, the root contains a list consisting of one image and one layout table. This layout table contains a link and an image as children. The transition between Figure 2a and Figure 2b shows how one can reorganize information within a PSG. For instance, if we want "link 1" to be another child of the root, we simply specify it as a child of the root, which also automatically deletes it from among the children of layout table 1.

Structuring Multiple Pages

Just as the PSG describes an ordering of objects resulting in a specific screen layout, the SSG describes an ordering of pages resulting in a specific navigational structure. While the kind of navigational links displayed may vary for each different site, the complete navigation structure is determined by the SSG. Also, the navigation scheme is completely unaffected by any structural changes once

Figure 2: A typical PSG. The transition between Figure 2a and Figure 2b illustrates how the structural layout of information on the webpage can be changed by an adjustment in the PSG.

that navigation scheme is specified for the website. Because our structure is guaranteed to always preserve meaning, it is impossible to "break links" or distort the structure in any way while updating or adding content.

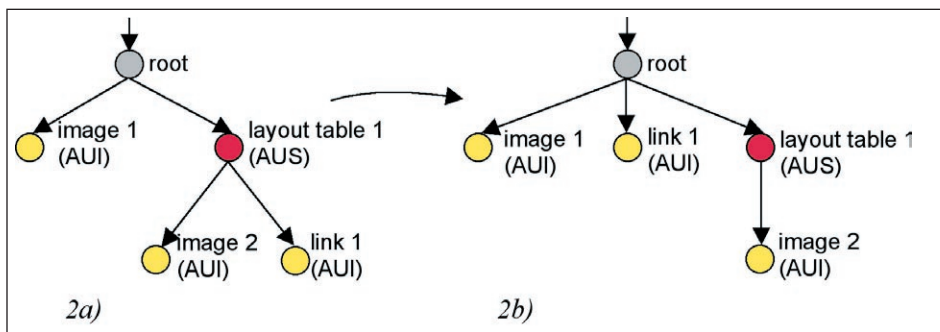The SSG is therefore constructed out of page nodes, each page being a PSG graph. Figure 3 illustrates an example of a small SSG. The root node contains some basic information about the page and some pictures and has two child pages, each with its own content. Also, "Page 4" of Figure 3 is not connected to the root and therefore not visible to website visitors. Nonetheless, "Page 4" can be made reachable from the navigational structure by specifying its parent page. Furthermore, one can also add part of the content of "Page 4" to any other PSG to make it "visible" within the navigational structure. Implementing an SSG as an acyclical graph rather than as a tree has some advantages. Since an acyclical graph allows disconnected parts, we can easily temporarily deny access to certain pages by not specifying its

parent without having to completely remove the page.

Another advantage of the SSG structure is that we can easily move objects across different web page boundaries rather than just within a single page. Just as a PSG remains valid if we change the ordering in the graph while obeying the afore-mentioned restrictions, the site structure also remains valid when moving objects between pages. Here, for example, we can move "image 3" into the root level of "Page 2" by incorporating it into the PSG that defines "Page 2." The same holds true for moving any piece of content, resulting in a site-wide separation of content, structure, navigation and design that is applicable to any web page.

Implementation

We chose to implement the ideas discussed above using a standard SQL database and readily available free software components. The code base was written in PHP (Version 5), and

we chose Apache and MySQL for hosting the data ([6], [7]). Our system is composed of two parts, the display engine and the content modification framework. The display engine receives the page requests sent by website visitors and generates the actual web page. In other words, the display engine queries the SQL database and builds an HTML representation from this retrieved information. The content modification framework allows the users of this CMS to change and update the web content.

The Display Engine

The display engine performs two distinct tasks by retrieving information from the SQL database. First, it queries the site structure graph to generate a navigation bar that is displayed according to a pre-defined template. Our implementation displays a navigation bar on the left side of a web page and displays the parent, children, and siblings of the current page in a hierarchical fashion. In addition, we utilize the same navigational structure information retrieved from the SSG to display a "depth indicator" at the top. The depth indicator shows all the ancestors of the current page, thereby allowing the user to go back to any parent page within the website (see Figure 4).

The second task performed by the display engine is to retrieve the content information of the current page from the PSG. As discussed before, this information retrieved from the PSG defines content as well as structure. We translate the retrieved information into HTML whenever a visitor requests a web page. Default header files and pre-defined style sheets which define default color, spacing, and alignment dictate how information is to be presented on the web page.
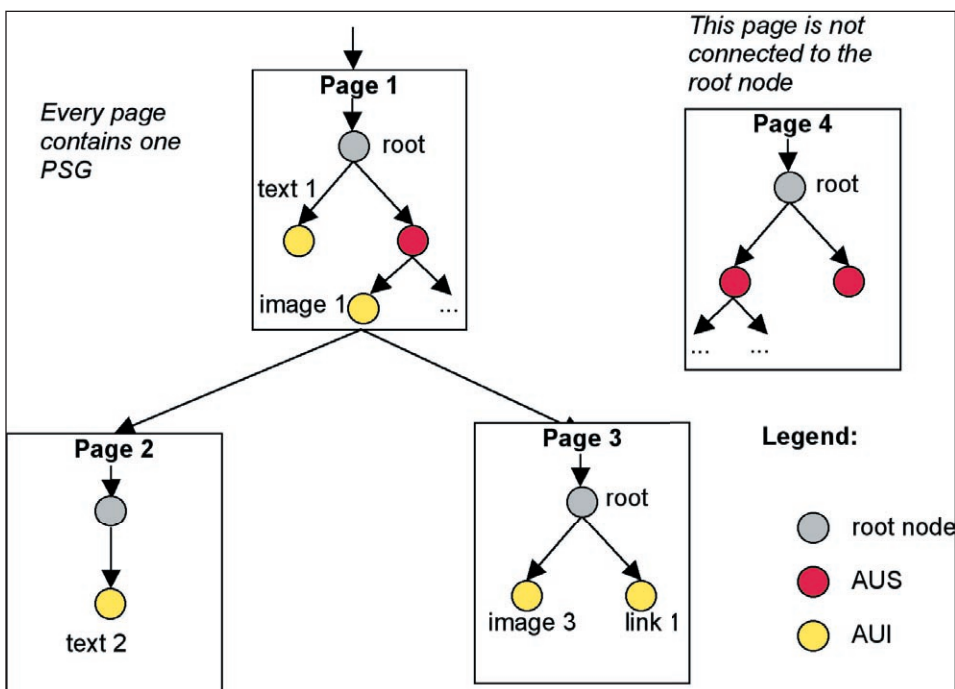
Figure 3: The relationship between an SSG and PSG's. Each PSG is a node of the SSG. The SSG defines the navigational structure of the website while the PSG represents the mapping between information and layout on any given page.

The Content Modification Framework

The content modification framework, here referred to as CMF, is a user-friendly environment that allows even non-technical users to easily update and add content. Because this framework is meant to be used by anyone operating the website to modify information, the ease of use as well as elimination of redundancy were the paramount design goals. The complete interface is accessed through web pages in order to ensure independence from the client's operating system. Access to this interface is restricted by username/password combinations.

After accessing the CMF, the user is presented with a list of pages currently available from within the site (see Figure 5). If the user wishes to create a new page, the parent of this page needs to be specified. A page is allowed to be deleted only if it has empty content and has no children in the SSG. With these stipulations, the SSG is always guaranteed to be well-formed without restricting flexibility and prevents accidental deletion of crucial pieces of information.

If an existing page is selected, all the content on the web page is displayed according to the structure

Here, it is illegal to move "layout table 1" into "layout table 2".
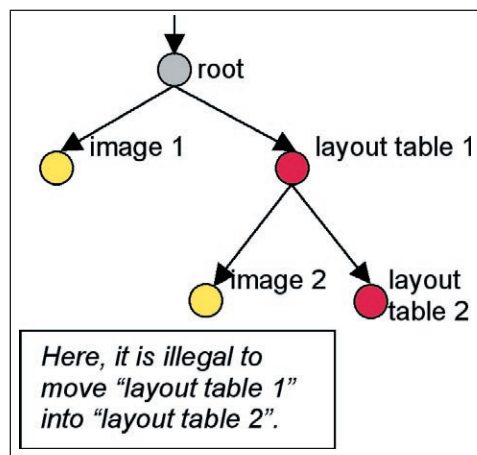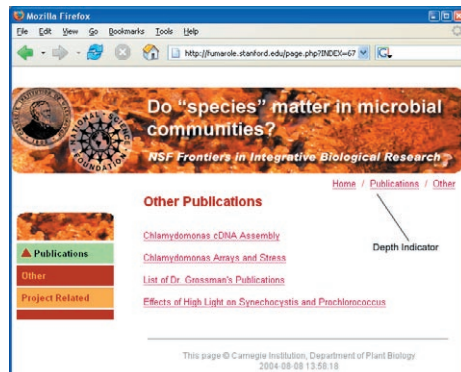
Figure 7

Figure 4: A page taken from the Carnegie Plant Biology website developed using the CMS discussed in this paper.

Figure 5: A user-friendly CMF shows all pages that are reachable in this website

indicated in the PSG so that the user can get a feel for what the real web page looks like. The user can easily create, modify, and delete AUI's and AUS's. It is also convenient to change the location of an AUI within the current page and to move it to a different page as discussed before (see Figure 6). Any allowed operation is guaranteed to result in a valid PSG. Graphically, moving content can be viewed as specifying a different parent for a given subtree of the PSG, where the parent can be an AUS within the current page or in a different page. Here, extensive checks are necessary to ensure the legality of a move operation. For example, it is illegal to move a node into some part of its own sub-
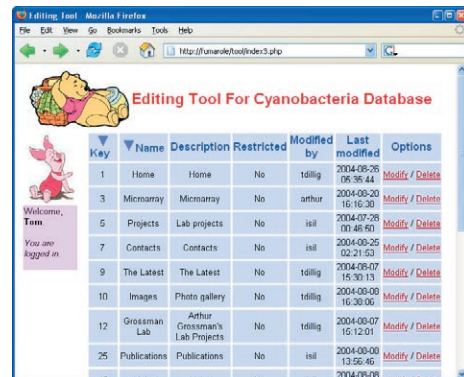
Figure 6: The content of one page, composed of one AUS (here Table1) and many AUI's. The structure of content represents the structure of the PSG.
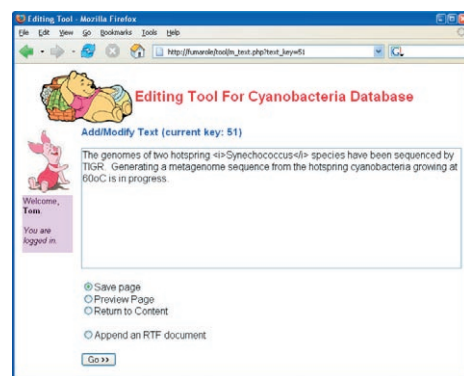
Figure 8: Text can either be manually entered or imported as an RTF document.
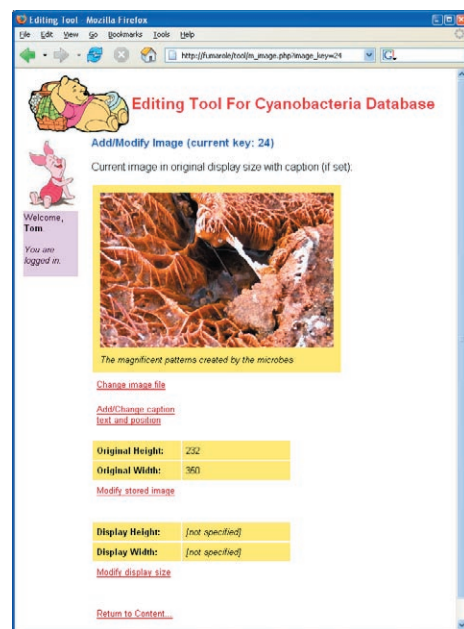
Figure 9: Images can be easily uploaded and resized without using external tools like Photoshop.

tree (see Figure 7). All the algorithms employed for this purpose are based on standard graph-traversal techniques and are a major source of complexity in implementing this stage. For an overview of the major algorithms modified for our purposes, see [8].

Every AUI and AUS can be added to the page currently under modification. Because AUI's always form leaf nodes in the PSG, only their content and parent needs to be specified when creating them; no layout information is necessary to fully define an AUS. For the AUI type "Text", the user can enter text with its correct HTML tags, such as those that define bold, italic, etc. (see Figure 8). In addition to specifying text using HTML tags, our system offers an RTF (rich text format, [9]) to text converter which automatically converts an RTF document to its HTML representation. The RTF format was chosen as a least common denominator, since all current text processing tools allow the user to save in RTF. The tool used in our implementation is a modified version of RTF to HTML, which discards any formatting applied to the RTF document and only extracts headings, paragraphs and bulleting/numbering [10]. Even though this process is based on heuristics, it does a reasonably good job when adding content to our reference page. Out of 17 documents added to our test site, all of which were at least a page long,

only six mistakes had to be corrected. While this fact was not established in a statistically significant way, the tool greatly reduces the effort and time required to publish existing Word/Text Processing documents as part of a web page—an extremely frequent need in many organizations.

Images are added in a similar fashion (see Figure 9). In addition to simply uploading the file, the user can also perform basic editing options such as resizing or turning. This feature, combined with a tool provided by our interface to resize every image to "reasonable" dimensions, makes the task of posting images easy and reliable for even the most inexperienced users. If a certain website features many pages with identical images, it will be more desirable to create an "image alias" pointing to one image to allow for easy duplication. However, the complexities introduced by this solution are real and its benefit is doubtful for sites benefiting most from this interface. Downloads and links are analogues to the AUI's already described; there are no special features associated with them.

When adding an AUS, the user is first asked to choose the dimensions of the layout table he/she wishes to use for content. At this stage, all the auxiliary attributes such as spacing and alignment are also specified. Then, already existing content on the current

page can be added to any field in the table if desired. The table size as well as all of its attributes and children can also be changed.

Conclusion

We described a content management system that can deal with any arbitrary structure and content while accomplishing the prime goals of any CMS: ease of content modification, separation of content and structure, and guaranteed well-formed pages. There are no restrictions on the kind of content that can be displayed. Even if it cannot be built form one of the described AUI's, the modularity of our solution allows new AUI's to be introduced very easily.

The most significant hurdle is performance. Every page construction by the display engine consists of multiple database queries and code to be executed. Even though we found this solution to scale extremely well to medium load, it cannot handle extremely high volume sites. This however can be easily compensated by caching. Many commercial and free tools exist for this purpose and a special version that regenerates pages only when the content has changed is in planning for our system [11].

Our system is currently deployed at the Carnegie Institute, Department of Plant Biology (http://fumarole.stanford.edu).

References

1. Dudek, David and Wieczorek, Heidi. "A simple web content management tool as the solution to a web site redesign." Proceedings of the 31st annual ACM SIGUCCS conference on User services, September 2003.
2. Sitecore .NET CMS. http://www.sitecore.net/.
3. The OpenCMS project. http://www.opencms.org/.
4. Vignette. http://www.vignette.com/.
5. Challenger, Jim, Iyengar, Arun and Witting, Karen. "A Publishing System for Efficiently Creating Dynamic Web Content." Proceedings of IEEE INFOCOM, 2000.
6. PHP – Hypertext Preprocessor. http://www.php.net/
7. MySQL – The World's Most Popular Open Source Database. http://www.mysql.com/.
8. Sedgewick, Robert. Algorithms in C++. Addison-Wesley Publishing, 1992.
9. RTFtoHTML. http://www.pa.msu.edu/reference/rtftohtml-docs/rtftohtml_overview.html/.
10. RTF Standard Version 1.6. http://msdn.microsoft.com/library/default.asp?url=/library/en-us/ dnrtfspec/html/rtfspec.asp/
11. K. Selcul Candan, Wen-Syan Li, Qiong Lup, Wang-Pin Hsiung Divyakant Agrawal. Enabling Dynamic Content Caching for Database-Driven Web Sites. ACM SIGMOD, May 2001.

Thomas Dillig and Isil Ozgener

Thomas Dillig is a junior, who is majoring in computer science and minoring in Management Science and Engineering. Tom was born in Munich, Germany.

Isil Ozgener is also a junior, majoring in computer science and minoring in mathematics. She is originally from Istanbul, Turkey.

Tom and Isil are currently both interested in areas of computer science related to program analysis and verification. They are both planning on pursuing doctoral degrees in Computer Science after getting their bachelor's degrees next year.

Mathematics

A Geometric Perspective on the Riemann Zeta Function's Partial Sums

Carl Erickson

The Riemann Zeta Function, $\zeta(s)$, is an important complex function whose behavior has implications for the distribution of the prime numbers among the natural numbers. Most notably, the still unsolved Riemann Hypothesis, which states that all non-trivial zeros of the zeta function have real part one-half, would imply the most regular distribution of primes possible in the context of current theory. The Riemann Zeta Function is the simplest of the Dirichlet series and is represented in its Dirichlet series form as $\zeta(s) = \sum_{n+1}^{\infty} n^{-s} = 1^{-s} + 2^{-s} + 3^{-s} + \dots$. This series only converges when the real part of s, $\Re(s)$, is greater than 1, outside the area of the complex plane relevant to the distribution of the primes. This area is called the critical strip: $\{s \in \mathbb{C} : 0 < \Re(s) < 1\}$. The result of our investigation of the geometric distribution will be to draw connections between the partial sums of the Dirichlet series and the value of $\zeta(s)$ with s in the critical strip despite the series' divergence. This article will illustrate connections between existing theory of the Riemann Zeta Function and geometric analysis of the partial sums through visual representations. From the connections between the visually accessible geometry and this theory, we illuminate and explore potentially provable improvements of the theory based on symmetry among the partial sums.

1. THE IMPORTANCE OF THE RIEMANN ZETA FUNCTION

Very complex mathematical ideas often spring from the investigation of questions that are simple to understand. The subject of this article - the behavior of the Riemann Zeta Function - is one such complex mathematical object. However, the study of the function's theory sprung from an investigation of the prime numbers' distribution. The prime numbers will seem very removed from most of the discussion in this article. 2, 3, 5, 7, 11, 13... they appear so simple at first. And the question of their distribution? Simply finding the next known prime might appear to be sufficient, but this is not the case. Mathematicians are unsatisfied with knowing simply a large number of primes and observing their distribution, for there are an infinite number of primes[1] [5]; instead, they seek a general rule that will dictate the distribution of the primes of any magnitude. This search gradually led mathematicians like Bernhard Riemann to utilize the theory of complex functions to describe the distribution of the primes [7].

Before discussing Riemann's work more, preliminaries about complex numbers and complex function must be laid out. Also, the experienced math reader should keep in mind that a rigorous exposition and careful attention to detail is

beyond the scope of this journal; the focus is to illustrate a connection between geometry and analysis.

Complex Numbers

Complex numbers are less familiar than prime numbers. Yet most of the numbers discussed, like the argument of the function s, the value of the function $\zeta(s)$, and each partial sum, are complex numbers. Thus, if complex numbers are unfamiliar, read this section before moving on or see [1] or more briefly [11].

A complex number is a number $a + bi$ where a and b are real numbers ($a, b \in \mathbb{R}$) and $i = \sqrt{-1}$. By high school we learn that -1 has no square root, which is correct considering only the real numbers, for no real number squared is negative. Complex numbers are critical objects exactly because they solve this problem: given any complex polynomial equation, all of its solutions are complex.[2] So, for example, the polynomial equation $x^2 + 1 = 0$ has solutions i and $-i$, which are complex numbers.

Geometric Interpretation of Complex Numbers

Whether or not complex numbers are new to the reader, it will be convenient to simply visualize of the complex numbers as a plane of numbers of the form $z = (a, b)$ where a is the

---

[1]Euclid proved that there are an infinite number of primes around 300 B.C.

[2]This property is called algebraic closure.

real part of z ($a = \Re(z)$), its component along the real axis, and b is the imaginary part of z ($b = \Im(z)$), its component along the imaginary axis. This plane is called the complex plane, and is shown in figure 1.

The "polar form" of complex numbers is important to their visual, geometric interpretation: any complex number can be represented by its non-negative distance from the origin r and its angle from the positive real axis θ as in figure 2. It is then equal to $re^{i\theta}$; an important geometric corollary to this fact is that multiplication by $e^{i\theta}$ is equivalent to a rotation by θ around the origin.

A final important complex operation is that of conjugation; its geometric meaning is also important. The conjugate of s is denoted by $\overline{s}$ and the conjugate of $a + bi$ is $a - bi$; the sign of the imaginary part is changed, which is equivalent to a flip over the real axis. For example, $1 + i$ is flipped to $1 - i$. A critical fact about conjugation is that for any complex function,

$$(1.1) \qquad f(\overline{s}) = \overline{f(s)}.$$

Conventions Regarding the Zeta Function

For the remainder of this text, let s represent the argument of the zeta function and let $\sigma = \Re(s)$ and $t = \Im(s)$ so that $s = \sigma + it$. Assume t to be positive, which does not result in any loss of generality because $\zeta(\sigma + it) = \overline{\zeta(\sigma - it)}$. The function $\zeta(s)$ accepts a point on the plane, $s = (\sigma, t)$, and gives an output that is a point on the plane. Likewise, a partial sum is the sum of the two components of each summand. Also, let $P_s(n)$ represent the nth partial sum of the series with argument s. Keep in mind that addition of complex numbers follows the pattern of vector addition:

$$(1.2) \qquad \begin{aligned} P_s(n) &= \sum_{k=1}^{n} k^{-s} = \sum_{k=1}^{n} \left(\Re(k^{-s}), \Im(k^{-s}) \right) \\ &= \left(\sum_{k=1}^{n} \Re(k^{-s}), \sum_{k=1}^{n} \Im(k^{-s}) \right). \end{aligned}$$

For example, figure 3 shows the the first five[3] partial sums of the zeta function when $s = 1/2 + 500i$. Each partial sum n differs from partial sum $n-1$ by the n^{-s}, or in vector form by $\left(\Re(n^{-s}), \Im(n^{-s}) \right)$. This vector forms the nth branch in the "connect the dots" pattern of the partial sums.

Analytic Continuation

A final preliminary issue with repect to complex numbers to resolve is the fact that $\zeta(s)$ is well defined for s for which the Dirichlet series $\sum_{k=1}^{\infty} k^{-s}$ does not converge. Many analytic complex functions like $\zeta(s)$ have an analytic continuation that extends a primary definition to additional areas of the complex plane. In the case of $\zeta(s)$, its Dirichlet series is

---

[3]The 0th partial sum is 0.

defined for $\Re(s) > 1$, but additionally, $\zeta(s)$ has a unique value for all values of s by continuation. For more infomation see [1] or [9].
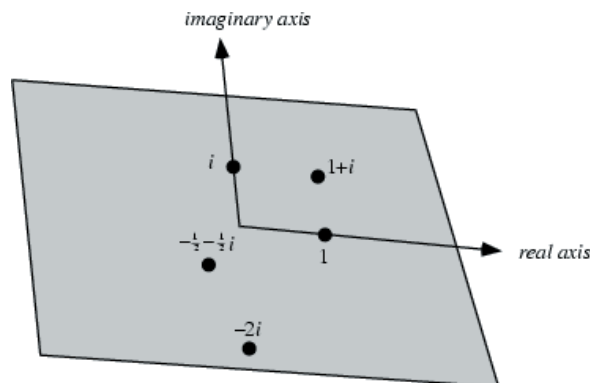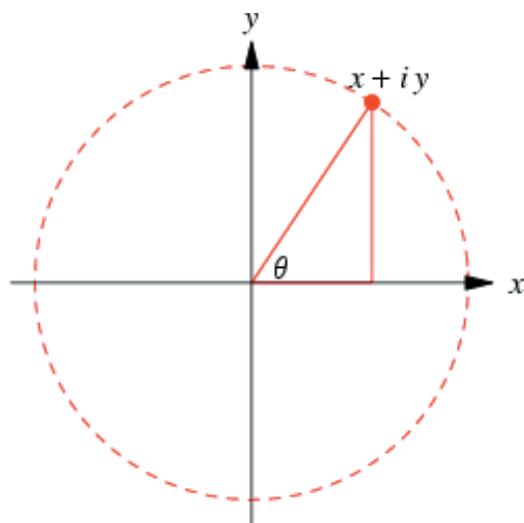

Figure 1: The complex plane. From [12].

Figure 2: A complex number $x+iy$ expressed in polar form, $re^{i\theta}$ where r is the circle's radius. From [11].
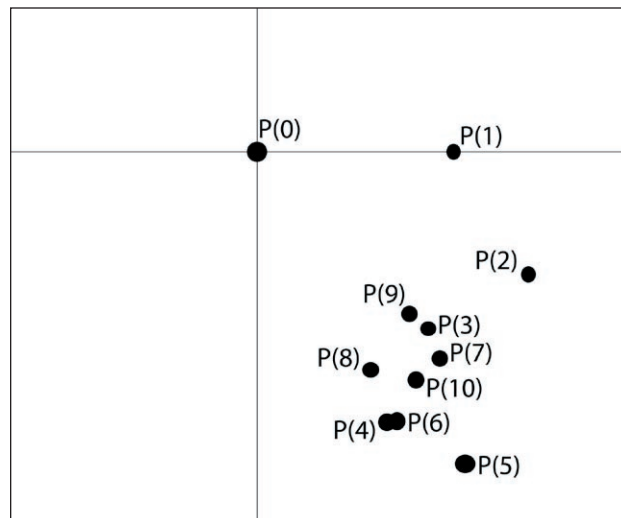

Figure 3: The first ten partial sums, $Ps(n)$, with $s = 1/2 + 500i$.

Riemann's Work

In 1859, Riemann proved a connection between the non-trivial[4] zeros of $\zeta(s) = \sum_{k \geq 1} k^{-s}$ and the density of the prime numbers among the natural numbers. In 1896, De la Vallée Poussin and Hadamard proved independently that the non-trivial zeros of $\zeta(s)$ lie in the critical strip (see figure 4), $0 < \sigma < 1$ (recalling that σ represents the real part of s), a fact from which the prime number theorem follows [8]. The prime number theorem is a rule for the density of the primes: it states that

$$(1.3) \qquad \pi(x) \approx \int_2^x \frac{1}{\ln x} dx$$

where $\pi(x)$ is the number of prime numbers less than than x [5].

Additionally, in Riemann's 1859 paper he stated a functional equation that implies symmetry across the line $\sigma = 1/2$: $\zeta(1-s) = \gamma(s)\zeta(s)$ where the gamma factor will be described later. Also, he formulated the famous Riemann Hypothesis, which conjectures that all non-trivial zeros lie on the critical line, $\sigma = 1/2$ [18]. The hypothesis, which states that the zeros of the zeta function lies on the critical line, implies that to the prime number theorem's estimate of the distribution of the prime numbers being as accurate as possible [8], [5]. Figure 5 shows the critical line in the complex plane along with trivial and non-trivial zeros.

The Behavior of the Zeta Function in the Critial Strip

To this day, the Riemann Hypothesis remains neither disproven nor proven and the behavior of $\zeta(s)$ in the critical strip remains mysterious; ten trillion zeros have been calculated, and all lie on the critical line [20]. See [3] for a detailed description of this mystery. The work I describe in this article may be the start of a new perspective from which to demystify the zeta function's behavior. Certainly, I do not claim any sort of significant progress toward the Riemann Hypothesis or important questions in analytic number theory; however, these methods may inform the analysis of the zeta function. Any conclusions that can be drawn about $\zeta(s)$ in the critical strip are significant, and these methods are first steps toward such conclusions.

2. OBSERVING GEOMETRIC PATTERNS IN THE PARTIAL SUMS

The Partial Sums and $\zeta(s)$

Figure 6 shows the the first 13000 partial sums of the Riemann Zeta Function $\zeta(s)$ with $s \approx 1/2 + 33704.56i$. The swirling shapes of the partial sums stand out in the figure. In fact, that shape is called a Cornu spiral (see figure 7).

---

[4]The Riemann Zeta Function has trivial zeros at all negative even integers [7].

While these spiraling shapes are significant, the first interesting pattern is not inherent in the picture itself: If a dot were put down for each partial sum after the 13000th one, the dots would not end at the final spiral, labeled C_0. They only get close to that point. But continuing onward, they fill up the picture spiraling outward in a circle. This final outward spiraling is a visual representation of the series' divergence: there is no point at which the series will end. This divergence follows from the fact that $1/2 + 33704.56i$ is in the critical strip, and there the Dirichlet series of $\zeta(s)$ does not converge.

The first key observation that began this research is that this final spiral appears to diverge in a regular, outward spiraling pattern around $\zeta(s)$ for all values of s in the critical strip. For example, C_0 is around 0 in figure 6, and $\zeta(1/2 + 33704.56) = 0$. Note that because the Dirichlet series does not converge for this point, $\zeta(s)$ is derived by analytic continuation and is not given by its series representation. Suspecting that this circling was no coincidence, I investigated a method for finding the complex number at C_0 with the hypothesis that any formula for it would be the value of the zeta function.
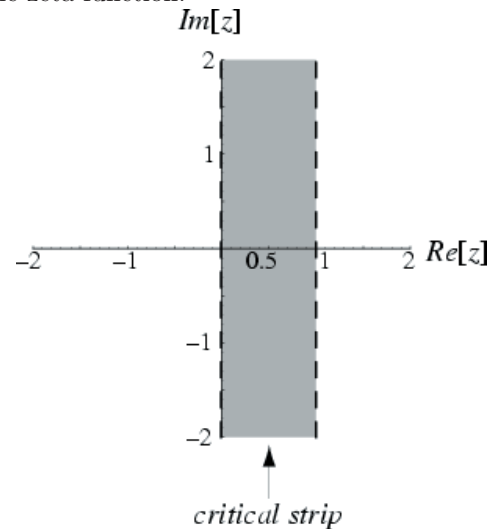
Figure 4: The critical strip: all of those complex number with real part greater than 0 and less than 1. From [15].
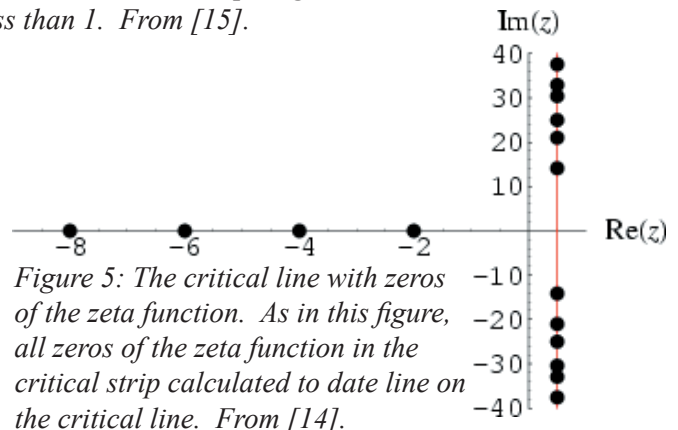
Figure 5: The critical line with zeros of the zeta function. As in this figure, all zeros of the zeta function in the critical strip calculated to date line on the critical line. From [14].

Figure 6: The first 13000 partial sums with s = 1/2 + 33704.56. This is a zero of the zeta function, evidenced by C_0, which is the same as $\zeta(s)$, being at the origin. Note the mirror symmetry between partial sums and the centers of the spirals across the line of symmetry.

Geometrically Finding the Center of the Spiral

The method to find this center consists of applying properties of smooth functions to three consecutive partial sums with the goal of finding a function to "correct" the outward spiraling back to its center. Properties of smooth functions can be applied to three consecutive partial sums because, in the critical strip, the distance between consecutive partial sums vanishes as the partial sum index increases. Visually, the partial sums become increasingly close together as their index increases, and begin to look like a smooth curve. Mathematically,

$$\lim_{n \to \infty} |n^{-s}| = 0.$$

Then, differential properties of curves can be expressed in terms of the partial sum index after being derived from the geometry of the partial sums. The experienced mathematical reader will notice that I discuss approximations without giving bounds on error. Unfortunately, this critical issue is beyond the scope of this article, but error bounds in central results will be stated in "big-O" form: $\mathcal{O}(f(x))$ denotes an error term bounded by a constant times $f(x)$ as $x \to \infty$ [10].
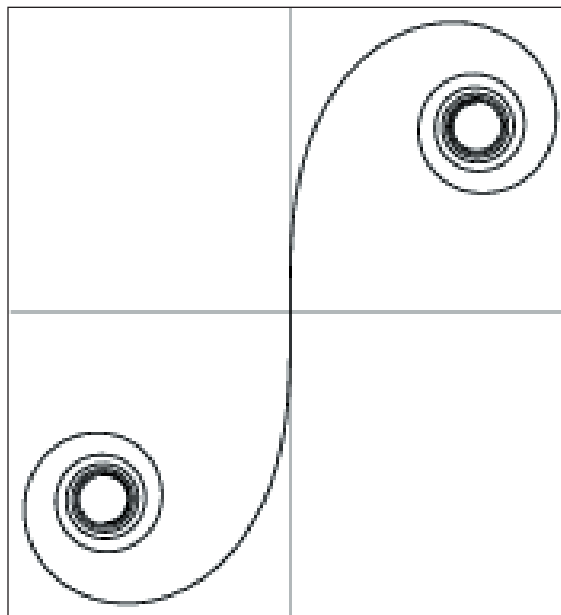
Figure 7: A cornu spiral. In a cornu spiral, the curvature is proportional to the distance along the curve from the origin. From [13].

25

Observe figure 8, showing nine partial sums from index $n+1$ to $n+4$ and their spiraling pattern. Because the angle change between the vector from $P_s(n-1)$ to $P_s(n)$ and the vector from $P_s(n)$ to $P_s(n+1)$ is approximately t/n and the length of each vector is $n^{-\sigma}$, the radius r_n of the uniquely defined circle going through partial sums $P_s(n-1)$, $P_s(n)$, and $P_s(n+1)$ (displayed in figure 8) is approximately

$$(2.1) \qquad r_n = \frac{n^{1-\sigma}}{2t}.$$

Recall that $s = \sigma + it$. Also, the length that the smooth curve covers as the index n increases by 1 is $n^{-\sigma}$ and the total length δ is approximately

$$\delta = \frac{n^{1-\sigma}}{1-\sigma}.$$

Combining these two facts, the following expression for curvature[5], κ, in terms of arc length δ follows:

$$(2.2) \qquad \kappa = \frac{1}{r_n} = \frac{t}{\delta(1-\sigma)}.$$

An equation expressing the curvature of a smooth curve in terms of arc length is called an "intrinsic equation" and determines the curve up to the factor of scale [2], [16]. This intrinsic equation is that of a logarithmic spiral (see figure 9). A logarithmic spiral is a polar function of the form $r(\theta) = ae^{b\theta}$. Solving this equation given the intrinsic equation, the b-constant is

$$(2.3) \qquad b = \frac{1-\sigma}{t}.$$

Then it remains to state the angle θ in terms of the partial sum index n and find the a-constant. A convenient fact is that the logarithmic spiral is also equiangular; that is, it intersects a given axis at the same angle each time it passes around. Therefore the tangent angle of the curve and the central angle θ differ by a constant, which is

$$(2.4) \qquad \arctan\left(\frac{t}{1-\sigma}\right).$$

The tangent angle in terms of the partial sum index n is

$$(2.5) \qquad -t\ln n.$$

With θ and b set, the scaling constant a is

$$(2.6) \qquad a = \frac{1}{\sqrt{(1-\sigma)^2 + t^2}}.$$

Combining all of these results, the function that shares its differential behavior with the partial sums is

$$(2.7) \qquad \frac{n^{1-\sigma}}{\sqrt{(1-\sigma)^2 + t^2}} e^{-it\ln n + i\arctan\frac{t}{1-\sigma}}.$$

Amazingly, recalling that $s = \sigma + it$, this function of n simplifies via complex arithmetic to

$$(2.8) \qquad \frac{n^{1-s}}{1-s}.$$

To check the equivalence between 2.7 and 2.8, apply the complex arithmetic rules from the introduction section.

Thus the center of this final spiral in the progression of the partial sums, which I label C_0, is given by

$$(2.9) \qquad C_0 := \lim_{n\to\infty}\left(\sum_{k=1}^{n} k^{-s} - \frac{n^{-s}}{1-s}\right).$$

It is necessary to take n to infinity because of the issues with error mentioned earlier. This error vanishes for large n.
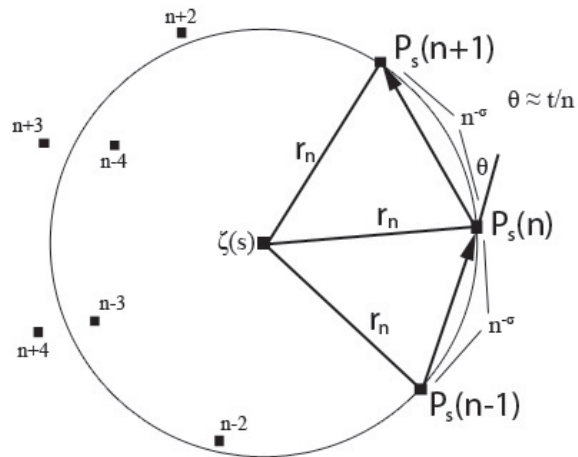
Figure 8: A demonstration of the geometric process for determining C_0. Given the lengths and angles, the curvature of the spiral and its center can be calculated. Keep in mind that this portion of the spiraling occurs as the partial sums spiral regularly away from C_0 after the point that they are displayed in Figure 6.
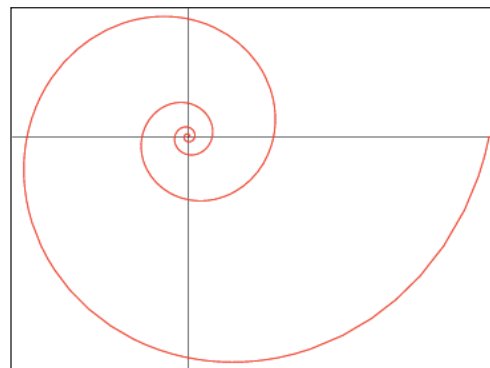
Figure 9: The logarithmic spiral. This is the family of curve that the partial sums approximate as they diverge in this spiraling pattern with C_0 in the center. From [17].

[5] The curvature of a smooth curve is the reciprocal of the radius of a tangent circle. Thus a curve with tangent circle of radius 2 has curvature of 1/2 and a straight line has curvature 0.

The Significance of this Result

Via geometric operations, we have found a center to the spiraling of the partial sums. By itself, the center has no meaning with respect to the zeta function. However, in 1921 the mathematicians Hardy and Littlewood proved by completely different means that this same formula is the value of $\zeta(s)$ [8], [4]. We may then conclude that

(2.10) $$\boxed{\zeta(s) = C_0}$$

This conclusion is the first key component of my research; however, the logical implictions are somewhat subtle. Note that nothing new has been proven about the actual value of the zeta function per se. Rather, we now know that the geometry of the partial sums is connected to the value $\zeta(s)$ in the pattern of divergence of the partial sums. Or in other words, the geometry of the partial sums has a connection to the value of $\zeta(s)$, giving an even finer characterization of $\zeta(s)$ than the Hardy-Littlewood identity alone implies. The patterns in the partial sums have been investigated by Hugh Montgomery and others (see [6]), but a direct geometric derivation of the connection between geometric properties of the partial sums (most importantly C_0) and $\zeta(s)$ is evidently new to this work. This new technique for approaching the zeta function may be useful, for any conclusions that can be drawn regarding the geometry of the partial sums are important; they may be used as a foundation for conclusions about C_0, and thereby $\zeta(s)$, even when s is in the critical strip.

3. SYMMETRY IN THE PARTIAL SUMS

The Approximate Functional Equation

We are now ready to discuss the main result of this article: the correspondence between geometric patterns and the approximate function equation for the Riemann Zeta Function. To do this, we will examine symmetries among partial sums for $\Re(s) = \frac{1}{2}$ that allow us to draw conclusions about C_0 that parallel the approximate functional equation. Then because we know from the previous section that $C_0 = \zeta(s)$, any fact about C_0 applies to $\zeta(s)$.

More precisely, the approximate functional equation is stated as follows.

Theorem 3.1 (Approximate Functional Equation). *Given $s \in \mathbb{C}$ in the critical strip $(0 < \Re(s) < 1)$ and real parameters $X, Y \geq 1$ such that $2\pi XY = \Im(s)$, then*

(3.1)
$$\zeta(s) = \sum_{k \leq X} k^{-s} + \gamma(1-s) \sum_{k \leq Y} k^{s-1}$$
$$+ \mathcal{O}(X^{1/2 - \Re(s)}(X^{-1/2} + Y^{-1/2}) \log XY).$$

Here $\gamma(s)$ denotes the arithmetic factor in the functional equation for the Riemann zeta function. It can be written as

$$\gamma(s) = \pi^{1/2 - s} \frac{\Gamma(s/2)}{\Gamma((1-s)/2)}.$$

This is an analog to the previously mentioned functional equation of Riemann, replacing the infinite series of the zeta function with finite sums and accounting for the error. Without a context, the consequences of this theorem are rather opaque; the geometric picture illuminates the statement and the structure that it describes. Consider figure 6, a example of the structure of the approximate functional equation. The figure shows the first partial sums of the Dirichlet series of the zeta function for $s \approx 1/2 + 33704.56i$. The part of the approximate function equation equal to these partial sums is the leftmost term, $\sum_{k \leq X} k^{-s}$. Each dot is a representation of this sum for successive X's - for example, $P(2)$ is this sum with $X = 2$. Now we will explain the rest of the approximate functional equation geometrically, referring to figure 6.

Understanding the Approximate Functional Equation Geometrically

Let us consider the simplest case for the value of Y: X and Y must be at least 1 and the second summand is taken over all integers less than or equal to Y, so let us consider $1 \leq Y < 2$. Then the sum $\sum_{k \leq Y} k^{-s} = 1^{-s} = 1$ and the value of the second term in the approximate functional equation is simply $\gamma(1-s)$.

To understand what values X takes on as Y varies from 1 to 2, the relation $2\pi XY = t$ is key; as Y varies from 1 to 2, X varies from $\frac{t}{2\pi}$ to $\frac{t}{4\pi}$ respectively. The approximate functional equation implies that all of these partial sums should approximate $\zeta(s) - \gamma(1-s)$, for the sum up to X is the only part of the equation that varies.

How can all of these partial sums of index in this range of X give approximations to $\zeta(s)$? The answer to this predicament lies in the Cornu spiral shape of the partial sums mentioned previously. We will investigate how the partial sum index X corresponds to the partial sum's position on the progression of spirals. Since we know that for very large X the partial sums spiral out centered at $\zeta(s)$, we can start by examining the index of the partial sums as they get closest to $\zeta(s)$. At this point, consecutive partial sums remain close to $\zeta(s)$ because consecutive differences between partial sums cancel each other out. This happens when the vector connecting $P_s(X-1)$ to $P_s(X)$ and the vector connecting $P_s(X)$ to $P_s(X+1)$ is near π; for then each difference between partial sums doubles back to where the last one started. Stated in terms of "connect the dots," if one drew an arrow from dot $X-1$ to dot X, and then from dot X to dot $X+1$ as in figure 8, the arrows would point opposite directions for partial

sums near one of the C_n in figure 6; we call these places C_n "centers."

We have seen that centers occur when consecutive partial sums double back on themselves. Now let's characterize this mathematically: th change in angle between these consecutive vectors in terms of the index X is

$$(3.2) \qquad \frac{t}{X},$$

which is an approximation for the difference of two consecutive terms of equation 2.5. Now we can turn this equation on its head to find at what partial sums these centers occur. Solving for X,

$$\frac{t}{X} \equiv \pi \mod 2\pi,$$

or, equivalently,

$$X = \frac{t}{\pi(2j-1)}$$

for some positive integer j. The mod 2π equivalence reflects the fact that the quantities are angles and all functions of angles are periodic with period 2π. We have now determined a family of X's for which $P_X(s)$ is near a center.

As labeled in figure 6, C_0 is the final center in the ordered progression of partial sums; therefore the index X corresponding to C_0 is thus the greatest value for which $t/X \equiv \pi$ mod 2π. This value is $X = \frac{t}{\pi}$. The index $X = \frac{t}{3\pi}$ is the second largest X for which the change in angle will be π, so the partial sums around this index must be at the second to last swirl - this swirl is labelled C_1 in figure 6. Each of the consecutive swirls obey this property as well, and each is labelled C_n leading up to C_0. Thus, in general, the index X of the partials sums around C_n is

$$(3.3) \qquad X = \frac{t}{\pi(2n+1)}.$$

Likewise, solving for the change in angle between consecutive partial sums congruent to 0 mod 2π, we may determine the indices X for which partial sums are halfway between swirls. By applying equation 3.2 to 0 instead of π, we find that the index of the partial sums halfway between C_n and C_{n-1} is

$$(3.4) \qquad X = \frac{t}{2n\pi}.$$

Parallels between Geometry and the Approximate Functional Equation

Now that we've found at what indices X the partial sums approximate the centers C_0, C_1, etc., let's use this knowledge to explain the relationship between X and Y in the approximate functional equation. Again, we will refer to figure 6.

As shown in the previous section, as Y varies from 1 to 2, X varies from $\frac{t}{4\pi}$ to $\frac{t}{2\pi}$. We now know that the partial sum with index $X = \frac{t}{3\pi}$ is located at the center C_1. According

to the identity $2\pi XY = t$, Y must be $3/2$ for $X = \frac{t}{3\pi}$. Therefore, the approximate functional equation claims, with appropriate error bounds, that

$$(3.5) \qquad \zeta(s) \approx P_s(\frac{t}{3\pi}) + \gamma(1-s).$$

This statement can be rephrased in terms of geometrically meaningful quantities:

$$(3.6) \qquad C_0 \approx C_1 + \gamma(1-s).$$

That is, the difference between the locations C_0 and C_1 is $\gamma(1-s)$!

To state other C_n in terms of $\gamma(1-s)$, let's consider other choices of Y and X in the approximate functional equation such that $Y < X$. When $2 \leq Y < 3$, then X ranges from $\frac{t}{4\pi}$ to $\frac{t}{6\pi}$ respectively, a of partial sums range on either side of the swirl C_2 around index $X = \frac{t}{5\pi}$. The resulting equation stated in terms of geometrical quantities is

$$(3.7) \qquad \zeta(s) = C_0 \approx C_2 + \gamma(1-s)\left(1^{s-1} + 2^{s-1}\right).$$

Extending this analysis in general for $Y = \frac{2m+1}{2}$ with m a positive integer, then $X = \frac{t}{m\pi}$ is the index of partial sums around C_m, and according to the approximate functional equation

$$(3.8)$$
$$C_0 \approx C_m + \gamma(1-s) \sum_{k=1}^{\frac{2m+1}{2}} k^{s-1} = C_m + \gamma(1-s)P_{1-s}(m).$$

Note that the sum with Y has argument $1-s$ instead of s as usual, and that summing up to $\frac{2m+1}{2}$ is the same as summing up to m.

This conclusion gives an approximate formula for the C_n:

$$(3.9) \qquad C_n = \zeta(s) - \gamma(1-s)P_{1-s}(n).$$

Now not only do we have simple approximation for C_n, but this equation suggests a correspondence among the partial sums of the Riemann Zeta Function. However, this correspondence is not direct; the factors that remain to be resolved are:

- The γ-factor.
- The relationship between $P_s(n)$ and $P_{1-s}(n)$.

Symmetry among the Partial Sums with s on the Critical Line

Both of these factors behave simply and imply a symmetry among the partial sums exactly when s is on the critical line; that is, the real part of s is one half: $\sigma = \frac{1}{2}$. For the remainder of this section, assume that $\Re(s) = \frac{1}{2}$.

The $\gamma(1-s)$ variable comes from the functional equation for $\zeta(s)$, which closely relates the value of $\zeta(s)$ and $\zeta(1-s)$. An expression for $\gamma(1-s)$ was given in the statement of

the approximate functional equation. The absolute value of $\gamma(1-s)$ is one if and only if s is on the critical line:
(3.10)

$$|\gamma(1-s)| = 1 \Longleftrightarrow 1 - \sigma = \frac{1}{2} \Longleftrightarrow \sigma = \frac{1}{2} \text{ (see [8]).}$$

Likewise, there is no simple relationship between $P_s(n)$ and $P_{1-s}(n)$ except when $\sigma = \frac{1}{2}$. When $\sigma = \frac{1}{2}$, then s and $1 - s$ are conjugate, for recalling the definition of conjugation we get that

(3.11) $\overline{s} = \overline{\sigma + it} = \frac{1}{2} - it = 1 - \left(\frac{1}{2} + it\right) = 1 - s.$

Any complex analytic function f satisfies $f(\overline{s}) = \overline{f(s)}$. Since a given partial sum is an analytic function of s, this powerful property ensures that

(3.12) $\qquad P_{1-s}(n) = \overline{P_s(n)} \text{ when } \sigma = \frac{1}{2}.$

These two facts together, $|\gamma(1-s)| = 1$ and $P_{1-s}(n) = \overline{P_s(n)}$, interpreted geometrically, imply a symmetry among the partial sums. First, note that because $|\gamma(1 - s)| = 1$, $\gamma(1 - s) = e^{i\theta}$ for some θ, using the polar form of complex numbers. Recall that multiplication by $e^{i\theta}$ is equivalent to rotation by θ and that conjugation is a flip over the real axis as we saw in the geometric interpretation of complex numbers.

Applying the new identity $\overline{s} = s - 1$ and keeping in mind the geometric interpretation of conjugation and multiplication by $\gamma(1 - s)$, then $P_{1-s}(n) = P_{\overline{s}}(n) = \overline{P_s(n)}$; with equation 3.9,
(3.13)

$$\boxed{C_n = \zeta(s) + \gamma(1-s)P_{1-s}(n) = \zeta(s) + e^{i\theta}\overline{P_s(n)}.}$$

The geomtric interpretation of this equation yields an algorithm to calculate C_n from $P_s(n)$:

- $\overline{P_s(n)}$: Flip the point $P_s(n)$ over the real axis
- $e^{i\theta}\overline{P_s(n)}$: Rotate the result by θ
- $\zeta(s) + e^{i\theta}\overline{P_s(n)}$: Translate the result by $\zeta(s)$

One may check that a composition of a reflection with translations and rotations is simply a reflection over a different line. Thus the approximate functional equation predicts a mirror symmetry over a certain line which is given by the factors $\gamma(s)$ and $\zeta(s)$. This symmetry is denoted by the line in figure 6.

The Implications of this Symmetry

The reader should keep in mind that this symmetry only holds for $\Re(s) = \frac{1}{2}$. Certainly, the conjugation functions and multiplication by $\gamma(s)$ have meaning when s is off the critical line, but recall that $|\gamma(s)| \neq 1$ for $\Re(s) \neq \frac{1}{2}$, so the

geometric correspondence between C_n and $P_s(n)$ involves a scaling factor, and hence destroys the symmetry. Perhaps the most important symmetry to keep in mind is that between C_0 and $P_s(0)$[6].

Now reach back to the beginning and recall the motivation for this entire discussion. Any sort of conclusion about the behavior of the Riemann Zeta Function in the critical strip is valuable, especially one that has possible relations to the Riemann Hypothesis, which conjectures that

(3.14) $\qquad \zeta(s) = 0 \Longrightarrow \Re(s) = \frac{1}{2}.$

Ideally, a symmetry result on C_0, which is identical to $\zeta(s)$, would show that the zeta function can be zero only when there is the prescribed symmetry among the partial sums. Since $P_s(0) = 0$ for any s, and there is only symmetry when $\sigma = \frac{1}{2}$, it may seem intuitively true that when $\zeta(s) = 0$, then there must be symmetry since $C_0 = P_s(0)$ and one point has mirror symmetry with itself.

Clearly this is not possible because huge approximations were taken along the way, and to boot, we defined the line of symmetry in terms of $\zeta(s)$, the quantity in question. There is a long and perhaps impassible path from any of these conclusions to any important conclusions. My work has a great deal to do with seeking well-defined notions of the C_n other than C_0.

However, this geometric approach is not without substance. The approximate functional equation only predicts an approximate symmetry between approximate locations of partial sums. It predicts nothing about the Cornu spiral shape! Of course, the proof of the approximate functional equation depends on this underlying truth that we can observe visually, but it does not capture the whole reality. This shortcoming suggests that the geometric interpretation of the zeta function's partial sums may have a surprisingly wide vocabulary to address the value of $\zeta(s)$ in the critical strip.

4. APPLICATIONS TO THE BEHAVIOR OF $\zeta(s)$ IN THE CRITICAL STRIP

Chronologically, my research progressed in the opposite direction of this exposition. After completing the work on finding that $\zeta(s) = C_0$, I visually observed the symmetry among the partial sums described in the previous section. It then took a good deal of investigation to find that this symmetry was predicted by existing theory, namely the approximate functional equation.

However, this fact does not mean that all geometric efforts are doomed to be fruitless because of their duplication of analytic results like the approximate functional equation. Though it goes beyond the scope of this journal, my work

---

[6]The 0th partial sum is defined to be 0, the sum of the first 0 terms

now consists in reproving the approximate functional equation using geometric techniques. This is surprisingly difficult given how "obvious" the approximate functional equation seems given the geometry. In other words, we can see visually that the approximate functional equation holds, so the geometric perspective on the partial sums and $\zeta(s)$ seems stronger than that of analysis. With this highly descriptive perspective, where we can "*see* the Riemann Hypothesis hold true," it would seem natural for a result like the less descriptive approximate functional equation would follow easily from the geometric perspective.

Though there is certainly no ease, there is some progress. Starting with the connection between the analytically meaningful $\zeta(s)$ and the geometrically meaningful C_0, my current work seeks to prove the approximate functional equation by proving equation 3.13, which is an expression for C_n, by induction, using a base case of $C_0 = \zeta(s)$ and then attempting

to show by geometrically analyzing the partial sums between C_n and C_{n+1} that the difference between the two is appropriate for equation 3.13. As of press time I have derived sufficient conditions in error bounds for results and am now working on those areas.

Ultimately, complete proofs having to do with the Riemann Zeta Function must use analysis, for geometry quickly falls into approximation. The usefulness of the geometric perspective on this problem lies in its applications to analysis. Research into smooth representations of the progression of the partial sums as in figure 6 suggest via the geometric perspective that variations on the methods of Fourier analysis, which unfortunately exceed the bounds of this article, may be sufficient to define the C_n clearly and investigate an exact symmetry. This method would provide a powerful tool for analyzing the Riemann Zeta Function in the critical strip.

5. Acknowledgements

References

[1] Ahlfors, Lars V. Complex Analysis, 3rd ed. New York: McGraw-Hill, 1979.

[2] Do Carmo, Manfredo P. Differential Geometry of Curves and Surfaces. Upper Saddle River, New Jersey: Prentice-Hall, 1976.

[3] Conrey, J. Brian. The Riemann Hypothesis. Notices of the American Mathematical Society 2003; 50, 341-353, http://www.ams.org/notices/200303/fea-conrey-web.pdf

[4] Hardy, G. H. and J. E. Littlewood. The Zeros of Riemann's Zeta Function on the Critical Line. Mathematische Zeitschrift 1921.

[5] Ingham, A. E. The Distribution of Prime Numbers. New York: Cambridge University Press, 1932.

[6] Montgomery, H. L. Ten lectures on the interface between analytic number theory and harmonic analysis. CBMS Regional Conference Series in Mathematics, Number 84. Providence, Rhode Island: American Mathematical Society, 1994.

[7] Patterson, S. J. An Introduction to the Theory of the Riemann Zeta-function. New York: Cambridge University Press, 1988.

[8] Titchmarsh, E. C. The Theory of the Riemann Zeta Function, 2nd ed. revised by D. R. Heath Brown. Oxford: Oxford Science Publications, 1986.

[9] Weisstein, Eric W. "Analytic Continuation." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/AnalyticContinuation.html.

[10] Weisstein, Eric W. "Asymptotic Notation." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/AsymptoticNotation.html.

[11] Weisstein, Eric W. "Complex Number." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/ComplexNumber.html.

[12] Weisstein, Eric W. "Complex Plane." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/ComplexPlane.html.

[13] Weisstein, Eric W. "Cornu Spiral." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/CornuSpiral.html.

[14] Weisstein, Eric W. "Critical Line." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/CriticalLine.html.

[15] Weisstein, Eric W. "Critical Strip." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/CriticalStrip.html.

[16] Weisstein, Eric W. "Intrinsic Equation." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/IntrinsicEquation.html.

[17] Weisstein, Eric W. "Logarithmic Spiral." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/LogarithmicSpiral.html.

[18] Weisstein, Eric W. "Riemann Hypothesis." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/RiemannHypothesis.html.

[19] Weisstein, Eric W. "The Riemann Zeta Function." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/RiemannZetaFunction.html.

[20] Weisstein, Eric W. "Riemann Zeta Function Zeros." From MathWorld–A Wolfram Web Resource. http://mathworld.wolfram.com/RiemannZetaFunctionZeros.html.

Carl Erickson

Carl Erickson is a sophomore from Milwaukee, Wisconsin. He is majoring in Math with his principal interest in number theory, and also dabbles in Computer Science and Religious Studies. He has interned with Lucent Technologies and will soon be researching algebraic number theory at Williams College and studying math abroad in Hungary. At Stanford, he works as a tutor and grader and has fun through racquet sports and musical pursuits. Carl would like to thank his advisor, Ben Brubaker, for his enthusiasm and support, and Professor Brent Sockness for his guidance in writing.

Druze and Jews

Adi Greif

How do ethnic groups construct their identity in terms of conflicts such as Civil Wars? Does the construction of identity shape conflicts? To help address these questions, the paper looks at the case study of the Druze heretical sect of Islam in Israel/Palestine before and during the 1948 War. The Druze over time cooperated more with the Jews and less with Muslim/native Christian factions. By what process did Druze choose to cooperate with Jews? My argument is that there were two simultaneous and interdependent processes: Druze were driven into an alliance with the Jews because of the downward-spiraling security dilemma between the Druze and Arabs, which was due to negative Arab reactions to Druze neutrality. Also, the Druze who collaborated with the Jews earned Jewish trust. Pro-Jewish Druze leaders and neutral Druze leaders first worked to ensure Druze neutrality, which engendered Arab hostility and the security dilemma.

When most people think about the "conflict" in Israel, they imagine fighting between two groups who have hated each other for decades: the Jews and the Arabs. This bipolar perception of the current fighting illustrates the victory of nationalist propaganda that has convinced the world that all the Arabs are one people, and all the Jews are one people[1]. A quick summary from the Jewish nationalist perspective of how Israel was created runs as follows: Jews, persecuted by everyone, fled to Palestine. The British, who controlled Palestine at the time, supported Jewish control of the area. When the British left, on May 15, 1948, the surrounding Arab countries attacked Israel. The Jews won this "War of Independence" and now control their state. The Arab nationalist story runs: Jews moved in, stealing land from Arabs. When the British left the Jews declared themselves a state and "ethnically cleansed" over a million Arabs,

around six hundred thousand of whom were forced to leave Palestine (Smith 2004, 200).

The problem with nationalist accounts is that that they assume the past existence of present nationalist identities, in this case that of two rigid identities: "Arabs" and "Jews" (Swedenburg, 1995). Yet these identities became more rigid over time; before 1948 it was not clear which communities would want to be considered "Arabs", nor which communities the "Jews" would consider "Arabs."[2]. When Israel became a state, an array of different communities existed, comprised of Sunni Muslims, Circassians, Ahmedans, Beduins, Greek Catholics, Greek Orthodox, Latins, Maronites, various Protestant sects, Armenians and Druze (Stendel 1973). According to a 1942 census, the Sunni Muslims were in the majority with 61.44% of the population, the Jews 29.90%, the Christians 7.85%. The remaining

1% were mostly Druze (Nissim 1980, 144). Of the minority groups who lived through the 1948 War, the Druze were unique in that they did not flee, and were not expelled (Parsons 2001, 68). This evidence suggests that by the 1948 War the Jews did not consider the Druze "Arabs" but instead viewed the Druze as allies of a unique status. In contrast to all the other communities, how did the Druze earn the trust of the Jews, and why did they choose to do so?

My argument is that there were two simultaneous and interdependent processes that eventually credibly committed the Druze to working with the Jews. One was that middle-level Druze leaders, after being contacted by Zionists who were interested in the potential strategic advantages of cooperation, earned the increasing trust of the Zionists through successfully accomplishing missions of tactical importance for them. Simultaneously, Druze

were driven into an alliance with the Jews because of the downward-spiraling security dilemma between the Druze and Arabs, which was due to negative Arab reactions to Druze neutrality. These two processes were interdependent because the Druze who collaborated with the Jews ("Druze collaborators"), along with neutral Druze leaders, worked to ensure Druze neutrality, which engendered Arab hostility and the security dilemma.

General Background to the 1948 War and Druze Involvement

Jewish immigrants began moving to Palestine in the 1880s, mostly fleeing intense Russian persecution. These early settlers arrived with a hodgepodge of motivations for what they wanted out of life in Palestine; over time, a nationalist Zionism that advocated a democratic and Jewish state became the dominant ideology (Prof. Zipperstein, class lecture 2004). For the land to be controlled by the Jews in a democracy, the Jews had to become the demographic majority. Thus, settlers adopted less inclusive policies toward locals (Shafir 1996, 235). Locals complained to the Ottoman government, which tried to set conditions on Jewish immigration. It, however, could not enforce its policy due to British influence and local corruption (Lesch 1979, 31-32).

After WWI, the British gained control of Palestine and (intentionally or unintentionally) furthered the Jewish cause through their policies. The first important articulated British policy was the Balfour Declaration, which Jews and Arabs interpreted as promising a homeland for the Jews, although the British intended for its

vague wording to promise nothing (Lesch 1979, 35-36). The British also decided to establish autonomous organizations for the Jews, Muslims, and Christians. The Jews most successfully took advantage of this policy to create administrative bodies, which they later used to run Israel (Lesch 1979, 79-80).

The perceived British focus on helping the Jews failed to resolve tensions between Jewish immigrants and Arab locals. Instead, British measures incited Arab rioting and Jewish terrorism. Notably, a large riot broke out in 1929, and the British responded with the 'White Paper' that limited Jewish immigration and land purchases. The British might have mollified the Arabs, but in doing so angered the Zionists. The Zionist-controlled Jewish Agency responded by creating the Joint Bureau for Arab Affairs to arrange contacts with Arabs and non-Jews such as the Druze (Parsons 2000, 20) and also formed the Haganah defense force. (Lesch 1979, 200-205). The last major revolt, from 1936-1939 by Arabs against the British, resulted in the destruction of any viable Arab leadership. Meanwhile, Jewish terrorist gangs attacked the British, worried that immigration restrictions would hinder the goal of a Jewish demographic majority (Smith 2004, 170).

The British asked the United Nations to propose a solution for the ongoing fighting, but when the U.N. proposed the "Partition Plan" to create one state for the Jews and one state for Arabs, the Arabs rejected the plan and started a "civil war" in 1947 (Smith 2004, 185). The surrounding Arab countries were only willing to help Arabs in Palestine if doing so suited their own purposes; Syria and Transjordan wanted control of the

Arab Palestine as demarcated by the Partition Plan. The day the British withdrew, the first Prime Minister of Israel, David Ben-Gurion, declared the existence of the state of Israel under the U.N. Partition Plan (Morris 2004 p 15). Syrian, Iraqi, Transjordanian and Egyptian forces immediately attacked Israeli forces, starting the "1948 War/War of Independence" (Smith 2004, 196).

Druze Background

The Druze were ostensibly good candidates for joining movements that resisted the Jews because of the Druze's cultural similarity to Muslims. The Druze are a heretical offshoot of Shiite Islam, founded in the 11th century. The Shiites soon forbade anyone to convert to the Druze religion so its practices and rituals became secretive. Only certain elite Druze even knew what the rituals were and so there are not necessarily salient religious differences between most Druze and the Muslims today (Betts 1998, 16). All Druze spoke Arabic and dressed like Sunni Muslims, although they did live mainly in separate towns or areas of towns (Betts 1998, 35-36).

Not only were the Druze culturally similar to the Local Arabs/Muslims, they were also similar in terms of political structure. Both societies were composed of family factions headed by esteemed religious leaders. The Palestinian Druze were heavily influenced by religious leaders living in Syria and Lebanon, notably the Syrian Druze al-Atrash family. The Sultan al-Atrash was famous for his revolt against Syrian leadership in 1925, for which he was exiled to Transjordan (Betts 1998, 88). Within Palestine, most Druze villages were clustered in the Northern Galilee,

with some in the Southern Galilee and Dalyat-el-Carmel areas (Parsons 2000, 18). It is important to note that the al-Tarif family dominated politics throughout all regions (Parsons 2000, 88). Other families that headed loyal factions tried to rival the Tarifs, notably, the Khayr family in the North and the Khunayfis family in the South Galilee (Parsons 2000, 18). For the sake of clarity, the dominant leaders such as the Tarifs will be called the 'elite' while the rival leaders will be called the 'secondary elite'.

The potential for a Druze-Arab alliance was never realized, and instead the Druze allied themselves with the Jews. The following case study will detail how this process occurred, providing empirical evidence for my argument.

Ussufiya: an Illustrative Example

Druze remained neutral in the early Arab riots for religious reasons, and this led a Jewish official named Ben Tzvi to initiate contacts that led to increased trust between Druze and Jews, as shown by Ben-Tzvi's shift from strategic to tactical justifications for an alliance. Druze leaders in 1929 sent a letter to the British proclaiming their neutrality on the grounds that the riot was a religious conflict between Jews and Muslims and so it did not concern the Druze (Firro 1999, 23). This action caught Ben Tzvi's attention because his job as co-director of the Joint Bureau for Arab Affairs was to create as many allies as possible, or at least minimize enemies. Ben-Tzvi, against the beliefs of his co-director, placed value on the strategic utility of the Druze. He thought that since the Druze in Syria, Lebanon, and Palestine all shared close bonds, befriending the Druze in Palestine would help the Jews influence Syrian and Lebanese

politics through the Druze that lived there, which was a Jewish policy goal (Parsons 2000, 21-22).

Ben Tzvi created contacts mainly among the secondary elite and continued to increase Jewish Agency trust in them through beneficial tactical exchanges. After failing to budge the neutral Tarifs, Ben-Tzvi approached the competing Khayr family, who befriended Tzvi in the hopes it would gain them power relative to the Tarifs. The Khayr lost the following political battle to the Tarifs and did not continue to play a political role until much later. After inter-family feuding, other rivals of the Tarifs[3] decided to help the Zionists (Firro 1999, 52-56). These Druze collaborators were also paid well for their services, not an insignificant factor in a then-stagnant economy. Tactical advantages included enforcing neutrality by convincing pro-Arab Druze not to aid the Arabs. For example, one secondary elite collaborator, Abu Ruqun, visited Lebanon to ask Druze not to help the Arabs, and to tell Beirut authorities to arrest Druze trying to enlist with the Arabs at the border (Atashe 1995, 41-43). Other Druze collaborators would meet with those trying to enlist from Lebanon and Syria and use various means such as bribery to convince them to return home (Atashe 1995, 57).

The actions of both the Druze collaborators and the neutral family leaders helped to rein in pro-Arab forces. The neutral Tarif maintained ties with the Islamic Supreme Council, and along with other neutral Druze leaders worked to convince the Pan-Arabic leader, Mufti Hajj Hamin al Husayni to put less pressure on the Druze to join resistance groups. Their argument was that the Druze in Palestine were weak, and thus the

Jews would crush them in retaliation if Druze joined Arab fighters (Firro 1999, 41). Druze collaborators used exactly the same argument to help convince the influential Sultan al-Atrash in Syria to keep Syrian Druze neutral (Parsons 2000, 31).

That neutral and pro-Jewish forces reined in the pro-Arab Druze forces angered Arab leaders who, in response, initiated a downward-spiraling security dilemma. During the 1936-39 Revolt, Palestinian Druze did not respond as a unit. The elite maintained neutrality and struggled to enforce it, but individuals nevertheless joined the Arabs. Some local leaders struck neutrality deals with Jewish authorities, and some were forced into organizing armed help for the Arabs. An illustrative example situation is that of the Druze town of Ussufiya. Ussufiya's villagers initially supported the Arab Abu Durra gang. After the gang severely mistreated villagers, murdering or abducting some of the leaders, the local elders decided to complain to the British. The British intervened and destroyed the gang, committing Ussufiya to help the British/Jews, as the Arabs would no longer trust them (Atashe 1995, 35-37). Only the previous year notables of Ussufiya had declined a British offer to collaborate, showing that Arab provocations had significant effect in Druze choice of action.

After the Druze betrayed the Abu Durra gang, more instances of Arabs targeting Druze occurred in Ussufiya and nearby towns, which convinced many Druze around Mt. Carmel that they must rely on local help, driving them into cooperating with the British and Jews, which in turn shut off alternative options. In 1938 Ussufiya residents sent a delegation to ask for compensation

and security from other Druze, the British consul, the nationalist Supreme Arab Committee, and the moderate Arab Nashashibi family faction, but despite promises of help the delegation received no aid. Sultan al Atrash, embittered that despite promises Arabs did not send aid during his rebellion against Syria, advised the delegation to depend solely on themselves for security (Atashe 1995, 49-52). Druze in the area organized local Druze self-defense forces that received arms from the British and sometimes coordinated with local Jewish forces also fighting Arabs (Atashe 1995, 55-57).

The case of Ussufiya illustrates the security dilemma as neatly defined by Posen: "This is the security dilemma: What one does to enhance one's own security causes reactions, that, in the end, can make one less secure (Posen 1993, 2)". The elite Druze chose to encourage neutrality in the 1929 riots due to security concerns and religious indifference; they felt they had no reason to endanger themselves by fighting and had no religious stake in the issue anyway. The actions taken by the neutral elite and the Druze-Jew collaborators to ensure neutrality led to an Arab view of the Druze as obstinately unhelpful, which led the Druze to decreased Druze security because Arabs such as the Abu Durra gang targeted them for attacks. The security dilemma situation at Ussufiya describes the start of what I term "downward-spiraling," because as a result of the situation just described, the security dilemma situation repeated and led to even more severe repercussions. Druze leaders in Ussufiya, to protect their town's security, betrayed Abu Durra to the British, but this led to increased Arab hatred of the Druze, which in turn held further security repercussions, which in turn led to more Druze-British/Jewish cooperation.

Synthesizing the Security Dilemma and Self-Interest Arguments

The Security Dilemma Argument explains that Druze-Arab relations constantly worsened through Arab hostility to Druze neutrality until the Druze had to befriend the British/Jews purely out of self-defense (Atashe 1995). The prime example of the security dilemma situation was the case of Ussufiya. The main problem with this argument is that it posits low Druze agency.

The Self-Interest Argument has a more activist focus than the Security Dilemma Theory; it posits that self-interested Druze leaders (secondary elite) collaborated with the Jews, and during the 1947 fighting these ties incrementally evolved into a Druze-Jewish relationship. It also shows that the Jews shifted from a strategy-based policy of contacting all minorities because of their connections in the surrounding countries to a tactical based policy of ensuring Druze neutrality or friendliness because of the military benefits (Parsons, 2000). The Khayr secondary elite family provides proof of self-interested Druze that adopted a pro-Jewish stance as a bid to gain supremacy over the Tarifs without resort to defense-oriented excuses. As already seen, Ben Tzvi shifted the Jewish Agency from a focus on strategic to tactical benefits. This theory is extremely strong. However, it posits that Druze-Jewish ties created out of self-interest lay dormant until the outbreak of civil war at which point the individual relationships evolved into an inter-community alliance; it also does not focus on security problems.

I synthesize the "Self-interest" and "Security Dilemma" arguments, with some modification. The problem with these two arguments is that, separately, they ignore the actors that the other argument focuses on. Together, they encompass the important actors. Parsons' "Self-interest" argument focuses on secondary elite collaborators but ignores leaders who were involved in promoting Druze security; this oversight ignores important cases such as that of Ussufiya. Atashe's "Security dilemma" argument focuses on leaders reacting defensively to Arab hostility, ignoring more activist, self-interested collaborators such as the Khayr family.

My synthesis includes all of the above actors: I agree with Parsons' argument that Jews, for strategic reasons, contacted Druze secondary elites, and over time and repeated instances of successful cooperation that increased trust saw the tactical advantages of a Druze alliance. My argument slightly modifies that of Parsons because she stresses that contacts became useful incidents only after the outbreak of fighting in 1947, while I posit a more gradual build-up of trust. Examples of these advantages being slowly built over time as seen in the case study were: exchanging favors (1929), convincing Syrian and Lebanese Druze to stay neutral (1936-1947), and even recruiting Druze fighters for the Jews (1948). Meanwhile, neutral Druze were driven away from Arabs and toward Jews because of the security dilemma. Resultant collaboration between Druze and Jews helped Druze prove themselves to the Jews but also shut off exit options with the Arabs such that both they and the Jews knew their only viable remaining choice

was collaboration. This modified Parsons argument can be synthesized with Atashe's argument because the secondary elite's collaborations helped drive the security dilemma. For the Jews, all the tactical benefits of Druze collaboration listed above helped ensure Druze neutrality, which in turn helped ensure Arabic hostility, setting up the security dilemma between Arabs and Druze. Druze collaborators, real or perceived, continued to upset Arabs, which in turn led to increasing the intensity of the security dilemma.

A problem with the two credible arguments and my synthesis is that any attempt to isolate important actors will inevitably lead to exceptions where other actors played important roles. The major example is the Syrian Druze battalion that eventually defected from the Arabs and fought with the Haganah in the 1948 War; the individual soldiers who defected out of self-interest before their commander defected are the important actors in this case. Yet, it is difficult to isolate patterns explaining how interactions led to trust without generalizing about actors; I decided to generalize.

Conclusion

Synthesizing these two theories more fully explains the processes leading up to increased trust: The Druze managed to convince the Jews to see them as separate from Arabs and credibly committed to the Jews for the same reasons the Druze chose to ally themselves with the Jews: because of 1) the downward spiraling security dilemma with the Arabs that drove them to eventually reliably collaborate with the Jews, and 2) relationships motivated by self interest that led to concrete benefits. From the Jewish perspective, a strategy-based validation for Druze cooperation slowly evolved into trust in terms of tactical operations through a series of credibility-building incidents.

For the Druze, a crucial first step was the neutral elite's role in ensuring neutrality and buffering Druze from Arab leaders, thereby allowing all Druze greater freedom to contemplate political choices. Secondary elite Druze could then help convince towns/Wahab's battalion to defect to the Jews for their own local reasons, including security against local Arab gangs. More collaboration, especially when it intensified the security dilemma with the Arabs, meant the Druze committed themselves more thoroughly to the Jews by shutting off exit options: they could not credibly return to the Arabs after betraying them and so were forced to continually betray them.

Policy Implications

The story of the Druze in Palestine before 1948 shows how the Arab-Israeli divide is historically contingent on the specific actions taken. That the Druze joined the Israelis and not the Arabs resulted from the specific nature of Jewish-Druze and Druze-Arab relationships and how they evolved over time into tactically motivated commitments. Members of other groups such as the Circassians and the Beduins also joined the Minorities Unit, but more of them left their homes during 1948. Presumably, the reasons why the Israelis trusted some of them and not others are as, one could say, banal, as why the Druze as a group were trusted – it depended on the relationships built and the interaction between elites, secondary elites, peasants, and leaders of other groups. The implications of this case study are that nationalist history tends to read history backward and solidify categories of identity, which prevents analyzing how minority groups can successfully avoid becoming labeled as "enemies" by a more powerful military group if they manage to construct their identity as separate from other "enemies" through credible commitments.

Notes

[1] This is a simplification for multiple reasons but still holds a lot of weight with many people and is thus interesting.

[2] For the sake of simplicity and brevity from now on I use the term I find most appropriate to convey my point: Jew, Zionist, Arab, or Muslim. The difficulty of successfully using non-nationalist terminology illustrates the success of nationalist rhetoric.

[3] For simplicity's sake families in this paper are associated with the most prominent and involved figure's political leanings. I also use last names and so ignore changes in actual characters, such as when a father dies and his son takes his place. However, families were not always unified in belief.

Bibliography

Al-Qattan, Nadja, Private Interview 2004.

Atashe, Zeidan. Druze and Jews in Israel – A Shared Destiny? Brighton: Sussex Academic Press, 1995.

Ben-Dor, Gabriel. The Druzes in Israel: A Political Study. Jerusalem: The Magnes Press, 1979.

Betts, Robert Brenton. The Druze. New Haven: Yale University Press, 1998.

Dana, Nissim. The Druse: A religious community in transition. Jerusalem: Turtledove Publishing, 1980.

Dana, Nissim. The Druze in the Middle East: their faith, leadership, identity, and status. Brighton: Sussex Academic Press, 2003.

Esco Foundation for Palestine. Palestine: A Study of Jewish, Arab, and British Policies. London: Kraus Reprint Co., 1947.

Firro, Kais M. The Druzes in the Jewish State: A Brief History. Leiden: Brill, 1999.

Kalyvas, Stathis. "The Ontology of "Political Violence:" Action and Identity in Civil Wars". Perspectives on Politics, 1:4, 2003, 475-494.

Khalaf, Issa. Politics in Palestine: Arab Factionalism and Social Disintegration, 1939-1948. New York: State University of New York Press, 1991.

Landau, Jacob M. The Arabs in Israel: A political study. London: Oxford University Press, 1969.

Lesch, Anne Mosely. Arab Politics in Palestine, 1917-1939. London, U.K.: Cornell University Press. 1979.

Morris, Benny. The Birth of the Palestinian Refugee Problem Revisited. Cambridge, Cambridge University Press, 2004.

Nisan, Mordechai. Minorities in the Middle East, second edition. Jefferson: Mc Farland and Company Inc., 2002.

Parsons, Laila. The Druze between Palestine and Israel, 1947-1949. Great Britain: Macmillan Press Ltd., 2000.

Porath, Y. The Palestinian Arab National Movement; Volume 2 1929-1939. New Jersey: Frank Cass, 1977.

Posen, Barry R. "The Security Dilemma and Ethnic Conflict".1993.

Shafir, Gershon. Zionism and Colonialism: A Comparative Approach. In Israel in Comparative Perspective. Ed. Michael N.Barnett. State University of New York Press, 1996.

Smith, Charles D. Palestine and the Arab-Israeli Conflict. Boston: Bedford/St. Martin's, 2004.

Stendel, Ori. The Minorities in Israel. Jerusalem: The Israel Economist, 1973.

Swedenburg, Ted. Memories of Revolt: The 1936-1939 Rebellion and the Palestinian National Past. University of Minnesota Press. 1995

Parsons, Laila. The Druze and the birth of Israel. Cambridge University Press. 2001, 60-79.

Zipperstein, Steven. Class lecture. 2004.

Adi Greif

Adi Greif is an undergraduate junior majoring in International Relations, with minors in Math and Middle Eastern Languages and Literature. Her research interests have ranged from anthropology to psychology to political science, and currently she is developing an interest in the Middle East and International Relations. This paper originated in a class taught by Professor James Fearon on ethnic violence. Her future research interests include an honors thesis on dialogue processes focused on Israel/Palestine and Ireland. Her research in dialogue groups grew out of her involvement as president of the Middle East Issues Dialogue Group at Stanford, where she leads discussions mainly between Muslims and Jews on campus. Her other activities include FUSION (Social Entrepreneurship), Israeli Folk Dancing. She has also been on the student leadership board of Hillel. Adi would like to thank Professor Fearon and Jesse Driscoll for their help with this paper.

Comparative Cultural Studies

"We're not gay; we're just foreign!"

Desi Drags, Disidentifications and Activist Film in New York

Ronak Kapadia

This piece considers a moment of South Asian queer cultural production in the diaspora, specifically the activist film "Julpari" made in New York City. The documentary, produced for the South Asian Lesbian and Gay Association (SALGA) by Khuragai Productions, follows a group of South Asian urban immigrant male drag queens as they build community, practice drag and complicate what it means to be an immigrant and queer in New York City. My work draws on the writings of performance theorist José Esteban Muñoz to consider the disidentificatory tactics these men employ in their everyday lives. Disidentifications elaborates on the quotidian survival strategies minority subjects use in order to navigate mainstream white and gay culture. Through an analysis of films, scholars can analyze how questions of belonging and desire are complicated for queer immigrant subjects who must consider how their non-normative desire and practices are imbricated with their identities as diasporic national subjects.

"Drag is not about life. It's about being larger than life. Always. … Hence, the eight inch heels!"
Faraz Ahmed in *Julpari*, Khuragai Productions

"Nina Chiffon," a.k.a. Faraz Ahmed, is a Muslim queer immigrant drag queen in New York City relocated from Pakistan. We first see her alone on a subway platform late at night, colorfully adorned in a green and pink sari, a tight black wig, and eight-inch heels to accentuate her already commanding height. Like any good New Yorker, she waits impatiently for the next subway train, her wig blowing gently in the wind. As the train approaches, Nina Chiffon scoops up her silk sari and disappears onto the train in an instant. Thus begins *Julpari*, a short documentary

produced by and for the South Asian Lesbian Gay Association (SALGA) in 1996. The film introduces us to Faraz Ahmed and his two queer desi[1] roommates Saleem and Saeed – who together call themselves a happy household of women. They are artists, performers and faithful members of SALGA – one of the most prominent South Asian queer organizations in New York City. *Julpari* features these South Asian queer men who speak openly about their lives – discussing, among other things: drag, family, belonging, and the politics of being queer and South Asian in America.

Swati Khurana and Leith Murgai, who together produced the short film, offer a valuable glance into a subcultural world easily marginalized within mainstream racist, heterosexist discourse and

representation. By affirming "a queer desi diasporic" social location, I argue these men complicate the various dominant accounts and existing assumptions surrounding mainstream queer and South Asian American identities. Through their recorded interviews and playful enactments of drag performance in *Julpari*, these subjects contest and subvert (even if temporarily) the hegemonic supremacy of heterosexuality and whiteness that has historically colluded to elide their existence.[2] While arguing that *Julpari* productively challenges representations of queerness and "South Asianness" throughout the film, this paper also exposes some of the contradictions and potentially conservative reinscriptions present in the film.[3] Ideally, these critiques will help sketch the politics of progressive

South Asian queer diasporic performance– a politics that speaks to the ways South Asian diasporic queers negotiate the specific and complex intersections of racial, gender and sexual power dynamics ordering their lives in contradictory ways.

Central to this effort is an acknowledgment that South Asian diasporic queer subjectivities enact contradictory politics. U.S. South Asian queers are not celebrated, exoticized or victimized in my reading of *Julpari* because I understand subjectivity to be complex and mediated through uneven relations of power. Thus, these individuals present both challenges and reinforcements to dominant understandings of desire, belonging, and citizenship. Throughout this analysis, I want to remind the reader of the history of "conservative impulses" within South Asian migrant communities in order to better contextualize the subcultural queer practices featured in the film. For instance, the continued pursuits of model minority status, upward technocratic/professional mobility, a lack of political organizing with other communities of color around affirmative action (only partially reconsidered post-9/11), and a history of generally privileged immigration for South Asian professionals to the US in the late twentieth century all complicate a critical reading of *Julpari*.[4] In addition, one can argue that entrenched heteronormativity and religious fundamentalism paralyze sections of the South Asian American community and aggressively police the expression of non-normative sexuality and desire. At the same time, however, the past fifteen to twenty years has seen a remarkable surge in progressive South Asian queer diasporic organizing – including organizations like SALGA and activist

films like *Julpari*.[5] I read this film in relation to both the "conservative impulses" and progressive activist histories within the larger South Asian diasporic community to illustrate how the conditions of diaspora and displacement, more generally, are not necessarily producing progressive (or neat) alternatives to static conceptions of "South Asian culture."

The following discussion investigates how the reproduction of dominant power relations develops in the film. I argue that *Julpari* and like forms of South Asian queer cultural production serve as a useful strategy of cultural criticism. These performances promptly intervene into the legacies of patriarchy, homophobia and fundamentalist nationalism attributed to the South Asian diasporic community. By sifting through both the transgressive and reactionary moments encoded in these works, one can enact a progressive politics steeped in feminist visions and queer perspectives on diasporic citizenship and belonging. What lies ahead is therefore an extended reading of specific moments in *Julpari* that deconstructs how these South Asian queers disidentify through drag and occupy space – subverting and inflecting dominant narratives of both queerness and South Asian migrant identity while also potentially exacerbating preexisting stereotypes.

Julpari begins with Faraz Ahmed, dressed in black, speaking to a camera that zooms into a close-up of his lightly powdered face.[6] We meet Faraz in mid-critique of the mainstream gay and lesbian community (read: white queers). He mentions that "being gay" is not just "that white buff pretty boy image" propagated by the media, asserting that being a person of color makes him more conscious of the racism

polarizing the queer community at large. Faraz's statement underscores the claim that South Asians and other third world migrants and people of color must protest their removal from queer representation – even by other queers. Faraz considers this dismissal "a crime" – shocked that other queers who have personally experienced dehumanizing discrimination at the hands of a homophobic populace would sanction further oppression. Clear from the start of *Julpari* is the fact that the directors were interested in portraying the friction existing between South Asian queers and the mainstream white queer community. They do so by centering on a migrant subcultural community that, until recently, was invisible in the "gay scene" of New York City - focusing exclusively on South Asian immigrant queers.

Later, the film documents SALGA's mini-vacation to a "Jewish village," of sorts. Saleem, Faraz's roommate, announces: "the whole SALGA gang was there" and ready to shock the townspeople by going out shopping "dressed in drag." One SALGA member, in response to the general alarm with which the whites greeted the South Asian queers, approached the camera and said: "We're not queer; we're just foreigners!" His statement light-heartedly riffs on the ironic and multiple positionality of the 'perpetually foreign' South Asian queer migrant who must simultaneously negotiate both the racism and the homophobia of the dominant population.

In order to further unpack this humorous political negotiation, I draw upon José Esteban Muñoz's theory of disidentifications, applying it to the diasporic context documented in *Julpari*. As a queer performance theorist, Muñoz is concerned with the

"world-making and world-mapping" potential of queer of color subjects. Muñoz defines disidentification as "descriptive of the survival strategies the minority subject practices in order to negotiate a phobic majoritarian public sphere that continuously elides or punishes the existences of subjects who do not conform to the phantasm of normative citizenship" (Muñoz 4). His work offers an imaginative and lucid framework to think through the ways desi drag queens both resist and reproduce dominant forms of representation that have absented or maligned their existence. Muñoz theorizes that queer performers of color can reformulate racist and homophobic images and symbols as potent counterhegemonic tools to destabilize power.[7] I argue that *Julpari* depicts non-conformist South Asian diasporic queers' "survival strategies" – most importantly through their use of drag – to negotiate this hostile public sphere. In fact, such works help minoritarian subjects like desi queers further articulate their collective protests against the racism and homophobia of their communities.

The film expands on its critique of normativity by documenting the lives of South Asian diasporic queer subjects. *Julpari* uses the SALGA retreat to offer its audience glimpses of how South Asian queers, in a sense, reterritorialize "the village." We see shots of the men dressing up in drag, acting out scenes from their favorite films, cooking together, playing ball outside on the lawn and running through fields hand in hand. One shot focuses specifically on a group of desi men, half naked, standing together under a water hose and waiting gleefully for the flash of a disposable camera to record the moment from the retreat. A water fight breaks out

in the midst of the pose, some of the members fall to the ground, and the ensuing force of the water douses the rest. This image of South Asian queerness has stayed with me long after watching *Julpari*. The ways these desi queer men occupy space together in this upstate New York villa offers an important window into understanding the subcultural practices of the community at large. Since cultural citizenship in the South Asian diasporic context (meaning an "authentic" sense of belonging to the nation or cultural group) is heavily invested in a specific heteronormative model of domesticity and family, *Julpari*'s content signals an important transgression against that norm. The film documents a group of single queer men cuddling, cooking and sharing bonds together – in a sense, re-imagining conceptions of kinship and family outside the constraints of heteronormativity. As one of the SALGA members states in the film, "Family is all about helping each other…I think SALGA fulfills that to a T." At best, connections to biological families are fraught and dubious for these subjects so it is important to consider how queer groups like SALGA – which alter the phobic landscapes of majoritarian public culture by advocating specifically for South Asian diasporic queers–serve as a meaningful family network. Indeed by claiming the resort village as their own, the SALGA members construct – at least temporarily - an ethnically queer counterpublic space where they can collectively contest the "hegemonic supremacy" of heteronormative family models, as Muñoz's writing on majoritarian public spheres predicts (Muñoz 1).

Beyond confrontations with ethnicity and public space, these desi drag queens actively 'queer'

representations of the South Asian diaspora. Saleem talks about how he has been "doing drag since [he] was seven or eight years old." He recounts a story of preparing for a drag performance as a child when his mother was away from home. Saleem would try on his mother's saris and then perform for the neighborhood children. When his mother arrived home early, as was often the case, he would simply explain that he was "just having fun, you know, it's like entertaining." Central to these men's narratives seems to be the elation and pleasure they experience in the production and performance of feminine beauty.

However, Faraz complicates that assertion by saying that "drag is not about wanting to be a woman. It is about these exaggerated forms of femininity and masculinity." I think Muñoz would agree, seeing male to female drag performance as a counteridentification with (meaning: a rejection of) female identities. Faraz extends his view on drag by saying that "it is about being in touch with 'qualities' … not necessarily identified as feminine or masculine … but qualities that are just … 'nice.'" *Julpari* depicts Faraz as wary of gendering his drag work, opting for a more fluid interpretation of gender roles.

Yet while confounding conventional narratives of gender and South Asian diasporic family networks through the very portrayal of non-normative desire, the queer agents in *Julpari* seem simultaneously complicit in the reinscription of a problematic gender binary. For instance, Faraz says "we view each other as sisters. Even though it may be considered misogynistic, we do view ourselves as a household of women." Although Faraz is conscious of a potential

masculinist reading of his statements, I still question the language he uses to name the experiential bonds he shares with these other desi queer men. I ask, why not a household of desi queer brothers? Or further, "a household of desi queens (non-descript gender)"?

Similarly, when asked to speak about Faraz, Saleem says: "Faraz? … uh, I don't know. *She's* like a wife. *He* cleans the place, and a lot of time *he* cooks. It's nice after a long day … I like to eat. I'm like a pig!" (emphasis mine). What is intriguing here is the ways in which Saleem ("the buffed up one" according to Faraz, remember) alternates between pronoun variants to describe his roommate. Also problematic is his patriarchal coding of cooking and cleaning as domestic spheres consigned to "wives," and thus, women. *Julpari* represents these desi queer males living together and documents the ways they make sense of their so-called "outlawed" existence as South Asian diasporic queers. Yet, their comments surrounding gender strike me as uncritical. Thus, although *Julpari* features South Asian queer diasporic subjectivities, they need not necessarily be politically libratory or more critical of static patriarchies that circulate within the larger diasporic community. Faraz is quick to condemn the white queer community for not witnessing the discrimination they participate in against queers of color, but he seemingly under-investiages his own South Asian male privilege. Missing, then, from *Julpari*, is a feminist critique of domesticity and labor.

While *Julpari* may offer a problematic portrayal of gender, more alarming perhaps is the apparent lack of racial critique in the actual desi drag performances documented in the film. *Julpari* centers on Nina Chiffon (a.k.a. Faraz) dressed as

Marilyn Monroe – mole and all - to a host of other feminine vixens, including the vamp Helen! The film crosscuts from images of Faraz as Nina Chiffon performing Marilyn Monroe: luxuriating in a black teddy, smoking a cigarette on a satin settee and seducing the camera to a Nina Chiffon redressed in a dominatrix corset lip-synching to Madonna's track "Vogue" and Sticky Finger's song "You Gotta Lick It, Before We Kick It." In the final moments of the film, we see her redone once more: this time she's sauntering through the subways of New York City in the green and pink silk sari and heels introduced at the start of this piece, gesturing longingly as if lost in the latest Bollywood film.

A possible concern here is that the high-femme drag performances and conflation of images depicted in *Julpari* elevate an image of beauty defined in terms of an imagined "global American otherness" – where Western feminine beauty aesthetics define the parameters of even South Asian drag performance.[8] This is alarming, of course, because, as this paper affirms, I consider desi drag (and its representation) urgent sites of cultural intervention in dominant homophobic and racist spheres. *Julpari*'s portrayal of desi drag seemed to endorse a white beauty aesthetic in performing femininity. It is worrisome that Faraz and his band of "sisters" were not effectively deploying queer desi performativity to imagine that diasporic queer politics envisioned at the start of this piece.

Muñoz's disidentification theory fortunately intervenes, lending a more complex interpretation. His text reminds us:

Identification … is never a simple project. Identifying with an object, person,

lifestyle, history, political ideology, religious orientation, and so on, means also simultaneously and partially counteridentifying, as well as only partially identifying, with different aspects of the social and psychic world. (Muñoz 8)

Extended further, Muñoz's theory claims that a queer drag queen of color like Nina Chiffon – while counteridentifying with a certain icon/ logic – perhaps Marilyn Monroe's "white high-femme" – cannot escape structures of power and dominance that inevitably locate Chiffon's South Asian queer male body within the performance and politics of drag. In essence, the drag queen can never imagine herself "outside of ideology" (12) because her body will always be situated against and within that ideology. That said, disidentification offers the drag subject a chance to rework the contradictory components of her identity together (i.e. sexuality, ethnicity, etc.) to both challenge and acknowledge dominant structures of power.

Even Faraz's drag name – 'Nina Chiffon' – can be read through this disidentificatory lens. At first glance, I was disturbed that Faraz's stage name did not have a "more desi sounding" resonance – like "Faraz Ahmed," for instance. Upon further observation, however, the queen's disidentificatory strategy becomes clearer. "Nina" has become a popular South Asian diasporic female name, as many first generation South Asians select baby names more palatable to an English tongue. Second, we realize that "Chiffon" is actually a popular fabric used in South Asian female clothing. Thus, a Pakistani immigrant queer male's appropriation of the name "Nina Chiffon" highlights the contradictory but creative ways

in which a queer performer of color authenticates this disidentification process within the discursive limits of dominant majoritarian spheres. Furthermore, even Khuragai's film name – *Julpari* - flares with disidentification. "*Julpari*" means "on fire" in Hindi, but has been recoded in desi queer vernacular to signify "flaming." However, "flaming," as we know, is a slang word originating in the white queer community – now apparently uprooted, translated, and deployed by South Asian queers in New York! Clearly, desi diasporic drag disidentifications must recycle, as they subvert, the language of both the dominant queer community and the mainstream South Asian community at large in a collective reformulation of identity.

To conclude, I will further emphasize the imperative project that disidentification sets out for the desi drag queens represented in *Julpari*. Muñoz warns:

> Let me be clear about one thing: disidentification is about cultural, material, and psychic survival. It is a response to state and global power apparatuses that employ systems of racial, sexual and national subjugation. These routinized protocols of subjugation are brutal and painful. Disidentification is about managing and negotiating historical trauma and systemic violence. ... I have wanted to posit that such processes of self-actualization come into discourse as a response to ideologies that discriminate against, demean, and attempt to destroy components of subjectivity that do not conform or respond to

narratives of universalization and normalization. (161) Muñoz's commentary reads especially alarming in light of recent increased repression of Arab and South Asian American communities in the U.S. after 9/11 and the continued denial of civil rights to immigrants and queers of all kind further sanctioned by the conservative policies of the federal government.[9] Queers of color are especially vulnerable to these types of authoritarianisms, making works like *Julpari* an explicit resistance to erasure.[10] As survival tactics, then, disidentifications offer South Asian queer subjects a space to examine and enhance their everyday lives. Cultural productions like drag performances often interrogate the complex ways in which queers of color are positioned in dominant culture. Muñoz's foreboding words speak to the ways South Asian queers have been rendered illegible from almost all forms of representation. Khurana and Murgai's *Julpari* begins to fill that void by offering one depiction of a South Asian diasporic queer community. Indeed, SALGA's film succeeds in portraying the ways desi queers refigure notions of family and community to account for their exiled subjectivity within mainstream spheres.

However, when Saleem happily recounts, "the whole SALGA gang was there" on their trip to the Jewish village, one must wonder who is being systematically elided even there. Where are SALGA's lesbians and drag kings, and why were they denied representation in *Julpari*? Where do the closeted second-generational SALGA members fit in? The bisexuals and transsexuals? And what about class, religious, caste, migrant and intra-ethnic inequities and allegiances? In critically evaluating

this film, I have wondered to whom and for whom *Julpari* and SALGA hope to speak.

Ultimately, these final questions may further the construction of and the discourse on the non-dominant, minoritarian communities discussed in these pages. It remains critical to consider who is written into these communities and who is simultaneously written out – thereby excised from the small repository of existing desi queer representation. Indeed, the impetus for this paper rests in the fact that South Asian queers are not well-represented in terms of visible cultural or academic production and their presence at the intersections of postcolonial and sexuality studies is still being worked out.

The desi progressive community needs more works like *Julpari* that will complicate the cinematic representations of our lives as South Asian diasporic people living in the U.S. These types of cultural works demand special attention to the ways we are positioned within and sometimes outside of discourse and representation. Muñoz's last words in his text speak to the larger project in which these South Asian queer films are involved. He writes simply: "our charge as spectators and actors is to continue disidentifying with this world until we achieve new ones" (200). The desi queers portrayed in *Julpari* are truly working within and beyond parameters of an unfinished project that seeks to imagine an egalitarian world where desi drag performance addresses and redresses a profound cultural critique that is more than life … and far more than eight-inch heels.

Bibliography

Bacchetta, Paola. 2002. "Rescaling Transnational 'Queerdom:' Lesbian and 'Lesbian' Identitary-Positionalities in Delhi in the 1980s." Antipode.

Cruz-Malavé and Martin F. Manalansan IV. 2002. "Introduction: Dissident Sexualities/Alternative Globalisms." In Queer Globalizations: Citizenship and the Afterlife of Colonialism, edited by Arnaldo Cruz-Malavé and Martin F. Manalansan IV. New York: New York University Press.

Gopinath, Gayatri. 2002. "Local Sites/Global Contexts: The Transnational Trajectories of Deepa Mehta's Fire." In Queer Globalizations: Citizenship and the Afterlife of Colonialism, edited by Arnaldo Cruz-Malavé and Martin F. Manalansan IV. New York: New York University Press.

Hawley, John C., ed. 2001. Post-colonial, Queer: Theoretical Intersections. New York: State University of New York Press.

Johnson, Mark. 1997. Beauty and Power: Transgendering and Cultural Transformation in the Southern Philippines. Oxford: Berg Publishing.

Julpari. 1996. Dir. Swati Khurana and Leith Murgai. Videocassette. Khuragai Productions.

Manalansan, Martin F., IV. 2003. Global Divas. Durham: Duke University Press.

Muñoz, José Esteban. 1999. Disidentifications: Queers of Color and the Performance of Politics. Minneapolis: University of Minnesota Press.

Prashad, Vijay. 2001. The Karma of Brown Folk. Minneapolis: University of Minnesota Press.

Puar, Jasbir K. 1998. "Transnational Sexualities: South Asian (Trans)nation(alisms) and Queer Diasporas." In Q&A: Queer in Asian America, edited by David L. Eng and Alice Y. Hom. Philadelphia: Temple Univ. Press.

Puar, Jasbir K. and Amit Rai. 2002. "Monster, Terrorist, Fag: The War on Terrorism & The Production of Docile Patriots." In Social Text 72 20 (3).

Footnotes

[1] The word desi refers to people who claim South Asian ancestry – including (but not limited to) Pakistanis, Indians, Sri Lankans, and Bangladeshis. While useful to politicize the term desi in order to reference the South Asian diaspora – it is important to critique the North Indian linguistic hegemony encoded in the term (it means "of country" in Sanskrit-based languages)– making it a term that is especially difficult to embrace by South Indians, or non-Indian South Asians.

[2] Notice here how this film was gendered. Representations of U.S. South Asian lesbianism or queer female to male drag are limited in *Julpari*. Indeed, claiming to address "desi queerness" at large without dealing with the gendered specificities of female queerness and drag performance would be unproductive and a disservice to the South Asian lesbian community – which, in addition to racism and homophobia, must also confront the sexist patriarchy of their own queer of color community. Thus, for the purpose of this essay, unless otherwise noted, "South Asian queers" will mainly speak for desi queer male to female drag performers.

[3] Through my reading of *Julpari*, I take as a given the fact that queer studies and postcolonial/ethnic studies have insufficiently investigated the ways sexuality and ethnicity impact each other in identity formation. See Jasbir Puar (1998), Martin Manalansan (2003) and José Muñoz (1999) for further discussion on the ways scholarship on queers of color importantly intervenes in the discrete disciplinary regimes that limit queer and ethnic studies, respectively.

[4] See Prashad (2001) for a lucid critique of South Asian American history and political organizing in the U.S.

[5] I don't mean to suggest here that groups like SALGA are automatically progressive or homogeneous. Indeed, some of these groups have vigorous internal struggles around racism, sexism and other issues that so-called "progressive" organizations encounter through their activism. By "progressive," I hope to index organizations and cultural production that organize around, and thus acknowledge, the multiple, intersecting sites of oppression impacting their community and those with which they share solidarity.

[6] References to *Julpari* assume citation to Swati Khurana and Leith Murgai's Khuragai Production, 1996.

[7] Muñoz's theory became useful to me in drafting this paper as I struggled to reconcile a perceived uncritical appropriation of the white beauty aesthetic of femininity by the desi drag queens in their performances. The conversation that unfolds here will hopefully unpack this dilemma further, clarifying how Muñoz's approach to drags of color performance helps to displace this tension.

[8] See Mark Johnson's *Beauty and Power* for his work on the gay/bantut in the Southern Philippines for further elaboration.

[9] This statement should not be read as suggesting that the repression of ethnic minorities, including Muslim, Arab and South Asian Americans is a new phenomenon post-9/11. Indeed, there is a long history of racialization of these groups in the U.S. and the "post-9/11" phrase certainly overdetermines the break from a previous period. My intention here is simply to notice the fact that specific racial and sexual discourses over the past three years have increased the production of a certain kind of racialized, sexualized subject as "terrorist."

[10] I'm thinking specifically here about how two distinct discourses have coalesced against immigrant queers of color. Notice how the 2004 presidential election was seen in the media as a referendum on "the war on terror" (e.g. increased surveillance/racial profiling, INS deportations and imprisonment of immigrant men) AND calls for "family values" (anti-gay marriage ballot initiatives, proposed homophobic federal legislation, denial of legal immigration status and services to HIV + immigrants, etc.). The combination of these racist, homophobic and xenophobic practices and proposals present specific challenges for immigrants queers of color. See Jasbir Puar and Amit Rai (2002) for further discussion.

Ronak Kapadia

Ronak K. Kapadia is a performer, writer and graduating senior at Stanford. He receives his degree, with honors, in Comparative Studies in Race and Ethnicity this June. He has been invited to present his research on campus and at conferences in Los Angeles, Miami, and Austin, Texas. Coupled with fieldwork experience in New York and New Delhi, Ronak has curated events for the Black Box Theater in Oakland and has acted in numerous plays at the college and regional level. Currently, he is working on a full-length solo performance piece entitled "Terrorist Fag Drag and the Wild, Wild West," a satire on post-9/11 racialization of U.S. Arab and South Asian communities. After graduation, Ronak will be in New York pursuing the Tom Ford Fellowship before enrolling in the doctoral program in the Department of Performance Studies at NYU'S Tisch School for the Arts.

Nanotechnology

Hydrogenation Energies and Vibration Frequencies of Hydrogenated Carbon Nanotubes

Exploring Possibilities for Hydrogen Storage

Ashok A. Kumar

Many properties and potential uses of carbon nanotubes (CNTs) have yet to be explored. Experiments have been conducted in some areas, such as hydrogen storage, but conflicting results were reported. This study adds to the basic knowledge of tube properties and provides justification for IR spectroscopy in studying hydrogenation energy. This study focused on the trends in hydrogenation and ground state vibration frequency of chemisorbed hydrogen, or hydrogen absorbed by chemical bonding, in zigzag-type carbon nanotubes. We used molecular dynamics (MD) simulations to determine the hydrogenation energy for several carbon nanotubes. We explored dependencies on tube size and hydrogen coupling, the interaction between two bonded hydrogen atoms. Results showed that frequency distributions found by IR spectroscopy could be useful in determining the distribution of sizes among a sample of carbon nanotubes. Larger nanotubes were found to be more favorable for hydrogenation, and hydrogen coupling created a propensity for hydrogenated carbon atoms to be separated by three bond-lengths. Additionally, we found that pairs of hydrogen atoms favored bonding along the cylindrical axis of a CNT over the equatorial direction. Alone, these results imply that hydrogen storage at high weight percents is not favorable. However, future research conducted on the possibilities of hydrogen bonded to the inside of the carbon nanotubes may prove otherwise.

Alternative energy sources are becoming increasingly important to investigate as world energy consumption rises and fossil fuels become depleted. Hydrogen is a strong candidate for a renewable, non-toxic, and plentiful energy source, and much research has been done to harness its energy [1, 2]. Yet, for hydrogen energy to become a reality, a suitable means for storage must be developed. Storage of hydrogen as a pressurized gas and cryogenic storage as a liquid can be dangerous [3]. Weight becomes an issue when considering storage in solid materials. The U.S. Department of Energy (DOE) has set a goal for hydrogen storage materials to have a fuel density of 6.5 mass % (mass % is a measure of the relative mass of the stored material to the total mass of the storage material) [3].

Carbon nanotubes (CNTs) have been proposed as a possible hydrogen storage material. With a high surface area to volume ratio, CNTs could store a high hydrogen content per volume if hydrogen atoms were to bond to the carbon atoms of CNTs. Carbon is also a relatively light element. If every carbon site in a single-walled CNT were hydrogenated, 7.7 mass% hydrogen storage could be achieved. However, great controversy surrounds the actual storage capability of CNTs. Experimental results vary greatly, with some groups claiming that hydrogen storage in CNTs is not a viable option while others remain optimistic that the high theoretical mass% estimates can be achieved [4,5]. Part of the problem in comparing experimental results to theoretical predictions lies in the current limitations of CNT fabrication.

What are Carbon Nanotubes?

Carbon nanotubes were discovered in 1991 by Dr. Sumio Iijima. They have a structure equivalent to a sheet of graphite rolled into a cylinder. Several possible configurations exist. The indices (n,m) are often used to distinguish types of CNTs. Using the reference frame of the adjoining figure, an (n,m) tube is one in which the unrolled unit cell has n hexagons in the A direction and m hexagons in the B direction. The zigzag studied here is of the type (n,0). The tubes can also be classified as single or multi-walled depending on how many layers of the tubular graphite structure are present.
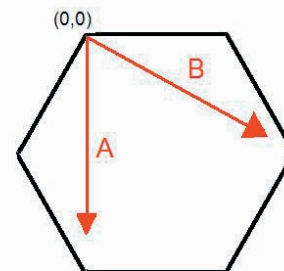
Experiments are often carried out on a mixture of single and multi-walled CNTs, many of which have different radii. Molecular dynamics (MD) simulations, which were used in this study, allow for a more complete understanding of experimental results.

Scientific Background

When a hydrogen atom binds to a carbon nanotube, it has, much like a spring, a certain resonant frequency associated with its bond energy, which is referred to in this paper as the ground state vibration frequency. Given thermal energy, the hydrogen atoms vibrate by varying bond lengths as a spring would vibrate if perturbed. Using MD simulations to find bond energies allows us to analyze whether this frequency is a function of size. Such a relation could allow experimentalists to determine how much of a sample of hydrogenated CNTs is of a particular size through IR spectroscopy. IR spectroscopy is already used to determine the amount of hydrogen bonded to the surfaces of CNT samples. Figure 1 depicts a hydrogenated CNT similar to those simulated.

By calculating hydrogenation energies, MD simulations can also shed light on how much hydrogen bonds to the CNTs and what configurations of hydrogen coverage are most favorable. The energy and ground state vibration

frequency of hydrogen atoms bonded to a CNT was found to be dependent on several aspects of the CNT. Correlations were found between radius, arrangement of surface hydrogens, and coverage.

Simulation Method

Use of MD Simulations

S. Park, et. al. used Vienna Ab-initio Simulation Package (VASP) to calculate the hydrogenation energy of carbon nanotubes of varying radius. CNTs of a larger radius were found to have higher, and thus less stable, hydrogenation energies than smaller CNTs [6]. Using a Brenner-Tersoff potential-based molecular dynamics

simulation, this trend was reproduced. As depicted in Figure 2, the MD simulations produce lower energies than are calculated by quantum simulations, but for larger radii, the trend is almost identical. The difference in values is to be expected. Studies by S. Park on the difference between Brenner-Tersoff potential and quantum simulations found a range of error of 0.61-0.67 eV or 21-40% for various carbon systems. The results of the CNT simulations, though slightly higher in their eV variation, were within less than 34% of the quantum simulation results. This combined with the fact that the trends are consistent between the two simulation methods allows molecular dynamics to be a useful tool in the qualitative
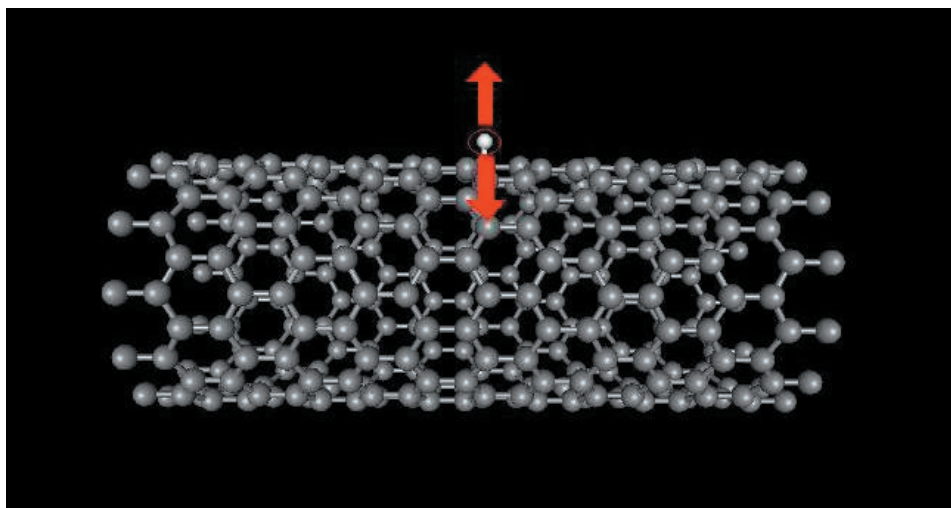
Figure 1: Model of hydrogenated CNT. Vibration frequency of the hydrogen was studied by varying the bond length and analyzing the potential energy.

study of the behavior of CNT's.

Simulations Varying Tube Size

To study the dependence of energy and ground state vibration frequency on tube size, we performed simulations on various sizes of zigzag type CNTs, both with a single hydrogen bonded to the surface and with no hydrogenation. Hydrogenation energies were determined by finding the total difference between the energy of the relaxed CNT-hydrogen system and the plain CNT system. Ground state frequency of the hydrogen was determined by plotting energy versus the varying length of the bond between the hydrogen and its nearest carbon. We fit a parabola to the data using the least squares method. Differentiating the parabolic equation then yielded a second degree ordinary differential equation that could be solved to find the frequency of the oscillations.

Simulations with Hydrogen Coupling

The hybridization of the carbon bonds also plays a significant role in the further hydrogenation of the carbon nanotubes. The carbon nanotube is deformed by the first hydrogen that bonds to it, significantly affecting the favorability of other sites for hydrogen bonding. Using a (12,0) tube, a second hydrogen was added in the simulations to produce two hydrogen configurations of the types shown in Figure 3. One hydrogen bonded carbon site was used as the reference, 0, denoted by the red dot. The second hydrogen atom was bonded to sites at various bond lengths from the reference. Both axial and equatorial orientations of the two bonded hydrogen atoms were simulated, where the axial direction is the one parallel to the tube's cylindrical axis and the equatorial direction is the one around the circular dimension
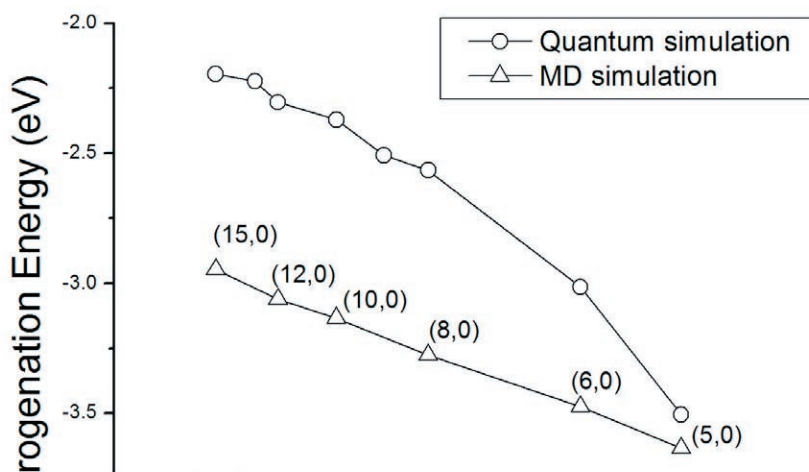
Figure 2: Using the initial pyramidal angle as an index of size, the relationship between size and hydrogenation energy show the same trend in both quantum and MD simulations.
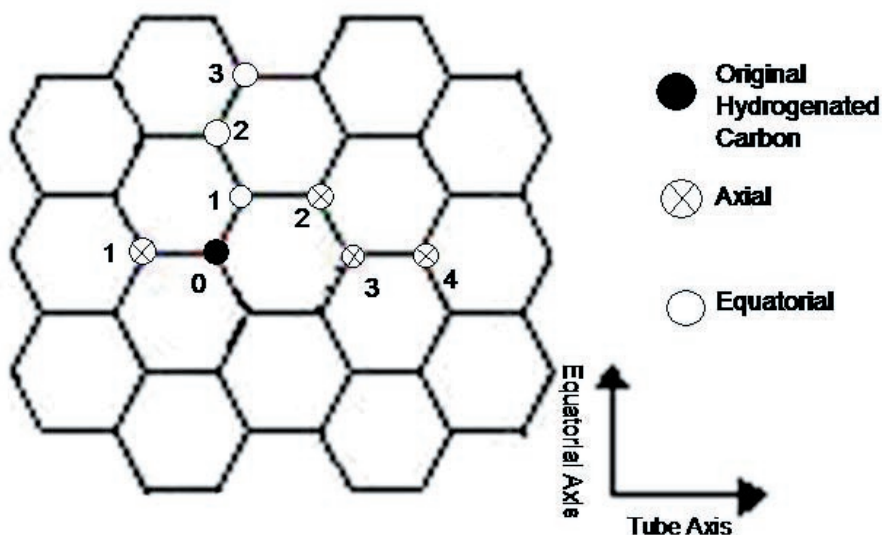
Figure 3: Various sites (blue and green) were studied for coupling effects between a second hydrogenated carbon and the original (red) carbon-hydrogen system.

of the cylinder.

Results and Analysis

CNT Size

The ground state frequency analysis from the simulation results with varying tube sizes are given in Figure 4. The hydrogenation energy of the carbon nanotubes is inversely related to the radius, while the ground state frequency shows a tendency to increase with radius. The rate at which the frequency changed as radius increased varied significantly, most notably between the (5,0) and the (6,0) nanotubes. This is most likely attributable to the nature of the carbon bonding. A carbon in a CNT
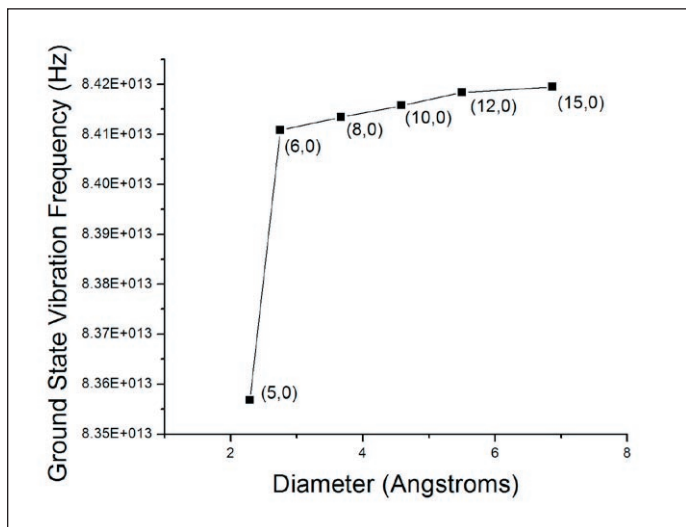
Figure 4: Ground state vibrational frequency has a positive dependence on size of the CNT.

Figure 5: Three bonds apart along the cylinders axis of symmetry is the most stable position for a second hydrogen atom. Axial separation was always at a lower energy and thus, more stable.

has a hybridized bond with characteristics of both the sp2 and sp3 bonds of carbon. In the (5,0) CNT, the hybridization is dominated by sp3 bonding while higher radii tend to have more sp2 character, like graphite. Since the type of bonding is so different, a larger difference would be expected between the (5,0) and (6,0) tubes than between other tubes that have similar sp2 hybridization.

Given the relation of frequency and radius, the frequency distribution measured through IR spectroscopy could be useful in determining the composition of a sample of carbon nanotubes. This would aid any process of separating or classifying tubes of a particular radius.

Figure 6: Frequencies along the tube axis are almost identical to the original vibration frequency for distances greater than 1 bond length.

Coupling Effect

Figures 5 and 6 summarize the results of this set of simulations. For the same bond distance, configurations of two hydrogen atoms along the two different axes show a tendency to favor the cylindrical axis. Furthermore, sites that are one or two bond lengths from the original hydrogenated carbon are much less favorable for a second hydrogen atom than carbon atoms that are further away. Carbon sites two bonds away from the previously hydrogenated carbon atoms are the least favorable for a second hydrogen atom. Once again, the frequency bears a slightly different dependency to the energy. The frequency at one carbon bond apart is drastically higher than the original vibration frequency, but once the hydrogen atom is placed three or more bond lengths apart, the frequency approaches the original value.

Conclusions

Analysis of the effect of hydrogen coupling on hydrogenation energy for the outer surface of CNTs demonstrates that high percent coverage is difficult to obtain. If only carbon atoms three bond lengths apart are hydrogenated, a maximum coverage

of only about 25% could be achieved before placing hydrogen on less stable carbons within three bond lengths of each other. This translates to only a 2.2 mass% storage of hydrogen. However, hydrogen could possibly bond to carbons from the inside of the tube and allow higher coverage.

Frequency results give little insight into the coupling effect but they are useful in assessing the size of CNTs. Since the simulations indicate that vibration frequency increases with size, IR spectroscopy could become a valuable tool for sorting various sizes of hydrogenated CNTs by finding the size distribution based on the distribution of vibration frequencies.

Acknowledgements

References

(1) Brian C. H. Steele, Angelika Heinzel. "Materials for fuel cell technologies." *Nature* Nov. 2001; 414, 345 – 352.

(2) Dane A. Boysen, Tetsuya Uda, Calum R. I. Chisholm, and Sossina M. Haile. "High performance solid acid fuel cells through humidity stabilization." *Science* 2004 303: 68-70.

(3) Schlapbach, L. and A. Züttel. "Hydrogen storage materials for mobile applications." *Nature* issue 6861 15 November 2001, Volume 414. Pages 234–378.

(4) Hirscher, M; Becher, M. "Hydrogen storage in carbon nanotubes." *Journal of Nanoscience and Nanotechnology*; Feb.-April 2003; vol.3, no.1-2, p.3-17.

(5) C. Liu, Y. Y. Fan, M. Liu, H. T. Cong, H. M. Cheng, and M. S. Dresselhaus. "Hydrogen storage in single-walled carbon nanotubes at room temperature." *Science* Nov. 1999; 286: 1127-1129.

(6) Park, S., D. Srivastava, and K. Cho. "Generalized chemical reactivity of curved surfaces: Carbon nanotubes." *NANO LETTERS*; Sept. 2003; v.3, no.9, p.1273-1277.

Ashok Kumar

Ashok Kumar (A.J.), was born in New Jersey on October 22, 1985. After being raised in South Dakota, he left home to attend Phillips Exeter Academy. He is currently a sophomore Physics major at Stanford University. He is the vice president of Phi Kappa Psi Fraternity and intends to continue studying nanotechnology over the coming summer.

Understanding the Market for Digital Music

Albert Lin

By 2008, digital music sales either as a-la-carte downloads or subscriptions are expected to reach $1.8 billion, up from $187 million in 2004—a ten-fold increase. Despite illegal file sharing activity, these figures indicate that the market for online music is rapidly growing. The success of the Apple iTunes Music Store suggests that consumers are willing to pay for online music that they can listen to anywhere and anytime. This paper begins by combining a discussion of the current market players and technologies with an analysis of relevant industry forecast data to understand what opportunities exist for current and future online music providers. Through pricing data research and consumer survey results, this paper argues that the present 99-cent per song strategy is optimal.

In 2002, digital music sales, which includes both paid downloads and online orders for hard copy music media, accounted for just over $1 billion. Informa Media forecast that by 2008 online music sales will amount to almost $3.9 billion. Purely online music distribution comprised of a-la-carte downloads and subscription services will make up $1.8 billion of all online sales in 2008, with the remaining consisting of online orders for physical CDs.

Though these future figures seem promising, the music industry remains concerned about piracy, pointing to the decline in retail music sales as evidence. Since peaking in 1999 with music sales of $14.5 billion, retail revenues have declined annually on average by 9 percent. In 2003, the Recording Industry Association of America (RIAA), an organized body representing several dozen record labels including the "Big 5" (Universal, Sony, EMI, Warner and BMG), reported a gross value of $11.9 billion shipments, a loss of 6 percent from 2002. Meanwhile, unit shipments of record media (CD, LP/EP, vinyl, DVD, music videos) have followed a similar decline from a high of almost 1.2 billion units in 1999 to just under 800 million in 2003.

Worldwide, the International Federation of Phonographic Industry (IFPI) estimated piracy reached $4.6 billion in 2002, an increase of 7 percent over 2001. However, because this figure neglects to include unauthorized sharing of music by private households, actual industry losses are even higher. Informa Media estimated losses of an additional $2.4 billion to the recording industry due to file sharing in 2004.

To curb the rise in file sharing practices by private consumers, the recording industry announced in 2003 that it would begin subpoenaing Internet Service Providers (ISPs) for consumer information in cases which the RIAA had reason to believe illegal file sharing had taken place. Given the individual's contact information, the RIAA would bring a lawsuit against the individual but would also be open to negotiations outside of court. As of July 2004, the RIAA has already subpoenaed information on nearly 3,000 individuals, and many private settlements have already been made on the average of $3,000. The effect of this unprecedented legal initiative is unclear as traffic on file sharing declined rapidly in the days and weeks following the announcement. But as of late, it has resumed the levels prior to the announcement.

In the face of these lawsuits, much industry attention is now focused on paid music services which offer users access to a wealth of songs for download at a reasonable price. The most prominent of these services
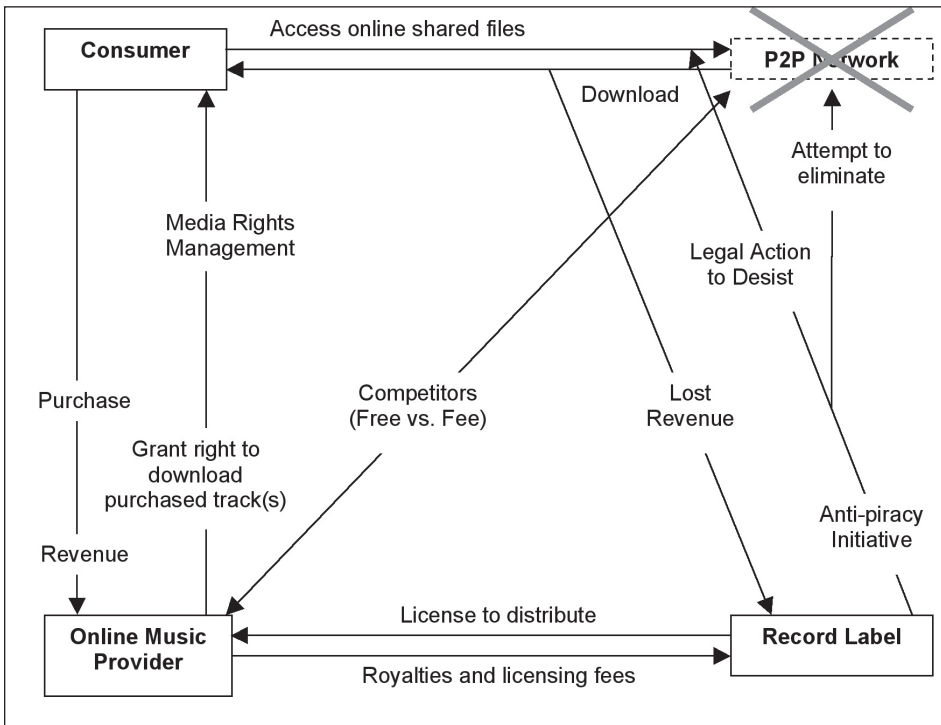
Figure 1

is the Apple iTunes Music Store (iTMS). With over one million songs to choose from and backed by the five major record labels, iTunes presents a tremendous opportunity to understand the digital music landscape. Unlike traditional markets, the key form of competition to paid online music services is not cost undercutting but reducing the incentive and opportunity for consumers to illegally obtain the same product for free through online file sharing networks. The same versatility that allows songs to be played through a consumer's home computer potentially allows the same song to be shared and obtained by millions of users worldwide.

The objective of this paper is to understand and capture the market opportunities that exist for digital music sales. I begin by exploring the key market players and their interactions. I also examine the problem of unauthorized file-swapping over peer-to-peer (P2P) networks and suggest that despite this activity, there still exists a market

for online music. Next, the paper follows with forecast data on digital music sales growth over the next few years. Having identified the present and upcoming prospects in digital music, I move to the core topic of analyzing a-la-carte download and subscription services to see how to most effectively price online music.

The Market Players

Like the typical market, goods flow from firms to consumers. For simplicity, I represent the supplying firms in the music industry as the record label. This is a good approximation because it is the record label which controls the licensing and distribution rights of an artist's work; the record label owns the copyright for the musical piece.

As indicated in Figure 1, goods flow from the record label to the consumer via an online music provider (OMP). Unlike traditional markets, no physical products are transferred; electronic payment and

download is the means of payment and distribution.

The process begins with the consumer who visits the web site of an OMP to browse and sample songs online. Payment is submitted electronically to the online music store and the user is allowed access to download the purchased songs. Using media rights management technologies, the music provider can allow and disallow permission to various end-user actions such as burning the songs to CD, creating duplicates, playing on different computers, and so forth.

The specific media management rights are determined by the record label which grants the OMP a license to distribute the songs. In return, the record label receives licensing fees from the OMP.

Also in the market presently is the recording industry's most targeted entity: P2P file sharing networks. A multitude of file sharing networks exist following the demise of Napster in 2001 which had an active user base of over four million users per day. In terms of music offerings on these networks, file sharing research firm BigChampagne reported that approximately 1 billion songs were available in June 2004 compared to the 820 million available a year ago. These networks allow users to acquire music digitally but without paying for the download.

The P2P Challenge

Competition in the digital music market is not so much cost under-cutting as it is determining how to limit free file trading networks. The ability for a consumer to acquire a paid product for free undermines the market for OMP and record labels

51

Table 1   U.S. Music Sales (In Millions)

	1997	1998	1999	2000	2001	2002	2003
Total Units	1,063.4	1,123.9	1,160.6	1,079.2	968.5	859.7	798.4
Total Value*	12,236.8	13,711.2	14,584.7	14,323.7	13,740.9	12,614.2	11,854.4

* Dollar value reflects the total suggested retail list prices of shipments.

Source: Recording Industry Association of America 2003 Year-end Report

alike. From the OMP standpoint, price competition is restricted given the existence of free versus fee; OMPs cannot offer the music for free because they must recoup the costs of licensing.

For record labels, it is not clear whether file sharing actually hurts music sales. The RIAA has long claimed that illegal file swapping has been a primary cause of the decline in music sales over the past five years (Table 1); they suggest that users who download music for free have little incentive to then purchase the CD.

However, contrary to the record industry's stance that P2P networks hurt sales, an empirical study by Oberholzer of Harvard Business School and Strumpf of UNC Chapel Hill in March 2004 found no statistically significant evidence that downloads lower music sales. Oberholzer and Strumpf examined data on actual file sharing activity over seventeen weeks in the last third of 2002. Focusing on U.S.-based users, they matched downloads to albums and then combined this with weekly music sales data to measure the effect of file sharing on sales.

The analysis by Oberholzer and Strumpf suggest that file sharing has a negligible effect on album sales. In fact, more successful albums actually benefit from file sharing—150 downloads leads to another album sale.[1]

If we accept the analysis done by Oberholzer and Strumpf, then the recording industry need not be too worried about file sharing. Moreover, for their most successful groups, file sharing should be more welcome than not. File sharing, despite ongoing legal efforts by the RIAA, has continued to thrive and evolve. The legal battle that brought the original Napster service to a close in 2001 saw countless new networks take its place, and as current services are shutdown, the file sharing underground will no doubt change and create more anonymous and decentralized distribution networks. The key for legitimate online music services is to differentiate themselves by providing exclusive quality content that users cannot readily find on P2P networks coupled with copyright policies that do not overly restrict music listening, burning, and portability.

Forecast

The online music market, despite file sharing concerns, is expected to grow to $3.9 billion by 2008, according to research firm Informa Media (Table 2). This figure includes both online sales of hard format media such as CD and vinyl, in addition to pure digital sales of music tracks and subscription-based

Table 2   Online Music Sales by Distribution Method ($ million)

services. Online sales as a proportion of overall music sales are forecast to reach almost 12 percent of the $32.5 billion music market in 2008, doubling the 5.3 percent figure in 2004. Pure digital sales of pay-per-download tracks and subscriptions are estimated to reach $1.8 billion, or equivalently 47 percent of total online sales in 2008.

Total online sales are expected to grow faster in the future as music providers improve features and usability, users become more accustomed to online music sales, and file usage policies adapt to consumer and industry compromise. Over 2002-2008, overall online music sales are forecast to grow annually at 22 percent. Pure digital sales from a-la-carte downloads and subscriptions are projected to increase 91 percent on average per year. A-la-carte download sales are expected overtake subscriptions in 2004 and account for over half of all pure digital sales in 2008.

By 2008, pure digital sales are expected to overtake hard format sales in North America with slightly over half (54 percent) of all online music sales consisting of a-la-carte downloads or subscriptions.

Globally, hard format sales will continue to dominate and hold a slight edge over pure digital sales at 52.6 percent. Nonetheless, by 2008,

Table 3   Online Music Sales in Top 10 Markets ($ million)

	2002	2003	2004	2005	2006	2007	2008
U.S.	419.1	470.4	553.9	712.6	964.5	1,265.5	1,696.5
Japan	203.5	233.0	257.8	295.5	385.6	478.8	602.7
U.K.	179.6	200.2	222.4	248.6	281.0	328.9	411.0
France	57.7	64.1	76.7	93.0	117.9	159.1	223.2
Germany	136.6	141.2	152.9	174.2	200.8	244.7	305.2
Canada	17.3	18.2	24.5	36.6	49.2	63.9	86.2
Italy	13.7	13.9	16.4	22.2	29.0	40.8	55.8
Spain	14.2	15.6	18.6	26.3	34.5	45.2	59.9
Australia	24.5	26.1	30.5	36.5	45.0	57.3	76.5
Mexico	5.2	8.0	10.2	13.0	15.9	20.7	27.0

Source: Informa Media Group (9/03)

Table 4   Top A-la-carte Download Music Providers

	Catalog	Track Price	File Format	Services				Usage Rights		
				D	S	FR	PR	L	B	P
Apple iTunes Music Store	1,000,000	$0.99	AAC, 128 kbps	✓		✓		5	7	✓[a]
Napster	700,000	$0.99	WMA, 128 kbps	✓	✓		✓	5	10	✓
Musicmatch Music Store	700,000	$0.99	WMA, 160 kbps	✓	✓	✓	✓	5[b]	7	✓
RealPlayer Music Store	600,000	$0.99; top ten tracks at $0.49	AAC, 192 kbps	✓		✓	✓	5	5	✓[c]
MSN Music	500,000	$0.99	WMA, 160 kbps	✓		✓	✓	5[b]	7	✓
			WMA							

digital downloads and subscriptions are forecast to represent a significant share of the market and likely to overtake hard format recordings in the years beyond.

In the top 10 markets for online music (Table 3), the United States is set to continue its lead as the largest market, accounting for 44% of all online sales globally in 2008.

Pricing Online Music

The record industry's fear of widespread piracy as a result of online distribution has resulted in restrictions on the use of paid online downloads. The fight against piracy, however, must seek a careful balance with consumer interest to freely use their purchased track just as if it had been purchased through a retail outlet.

This section first discusses the a-la-carte download and subscription-based pricing models and then analyzes various price levels against consumer willingness to pay. The objective here is to understand what fee structure is most profitable.

A-la-carte Downloads

A-la-carte download services like the iTMS charge users a set fee for each track or album downloaded. The flexibility for consumer choice in purchasing music makes a-la-carte download services appealing: users are free from the traditional "bundled" goods offered by the record labels.

As shown in Table 4, most a-la-carte download music providers offer individual songs at 99 cents per download; hence, much differentiation in this market centers on usage rights, catalog size, and exclusive content such as pre-releases or promotion tracks. Many of these OMPs also offer streaming content from radio stations and audio books for purchase.

Apple iTunes Music Store

The premier pay-per-download music provider is currently the iTMS. With the largest catalog size of any OMP, the iTMS has become nearly synonymous with paid online music: it commands a 70 percent market share in legitimately downloaded singles and albums, and in July 2004, it surpassed 100 million downloads.

Songs in the iTMS are offered using the AAC file format at 128 kbps, which is of higher quality than the MP3 format at the same bit rate. Moreover, unlike MP3s, they can be digitally protected from unauthorized re-distribution and re-play.

Apple's decision to adopt the AAC/FairPlay specification pushes users towards its iPod portable MP3 player, the only portable media device that can play the files purchased from the iTMS. Nonetheless, the iPod has become wildly popular in the United States capturing 58 percent of the market for portable MP3 players as of June 2004. According to Apple, the iTMS now sells over 16 million songs per month, or an equivalent 200 million per year, making it the most widely used paid music service.

Beyond its massive song catalog and easy-to-use interface, Apple differentiates itself with exclusive content, community tools, and one of the largest audiobook

collections. The iTMS features exclusive releases from many independent labels and also offers a free new single every week on Tuesday for download. Discovering new music is simple with Apple's extensive playlist browsing, sharing, and recommendation features. Lastly, the iTMS boasts one of the largest audiobook collections among its peers with over 5,000 titles available starting at $2.95 per item. Users can transfer these audiobooks—be it bestsellers, magazines, language books, and even public radio shows—for conveniently on-the-go listening using the Apple iPod.

Subscriptions

Just as consumers pay a monthly fee to access the Web through their Internet Service Provider (ISP), OMPs offering subscription-based services charge a monthly fee to access and download music. A listing of the top subscription services is shown in Table 5.

Napster

Napster's core product is the monthly subscription service which allows users to stream its 700,000+ song catalog and download songs for playback on up to three computers. Burning the songs to CD requires actually purchasing the tracks, of which subscribers can save up to 20% by buying multiple tracks at once. Napster subscribers also get access to over 50 commercial-free stations and the ability to build custom stations. Users can save songs to their music library for later playback.

In the beginning of September 2004, Napster launched a beta of its Napster-To-Go service, which in addition to the regular subscription service also allows unlimited transfers to a portable media device—so long as the subscription remains active. Previously, Napster's subscription service did not allow transfers to portable players. Preliminary pricing has been set at $14.95 per month.

Despite all of Napster's new features and shift towards becoming a legitimate paid service, it still retains the same tight community focus as before. While it no longer offers easy file sharing access, Napster does allow users with similar music tastes to browse and copy playlists and even see what all of Napster is listening to in real time. Users can send track and playlist recommendations to other Napster and non-Napster users using the provided Napster Inbox, a feature unique to this service.

Pricing Analysis

The popularity of Apple's iTMS suggests that its pricing structure must have at least some appeal to consumers, who can now find many credible alternatives such as Napster, Musicmatch and MSN Music. While OMPs such as RealNetworks and Wal-Mart have tried to undercut the digital music industry's predominant "99 cents per track" standard, they have found it difficult to maintain consistent profit margins in an already razor thin margin sector. This section analyzes various pricing levels by relating them to both consumer willingness to pay and revenues.

In 2003, a survey from Jupiter and Ipsos-Insight revealed that the standard 99-cent price being offered by OMPs was the most preferred, assuming that the songs purchased could be freely copied and burned to CD. These results are reflected in Table 6. Young adults, or those ages 18 to 24, preferred a lower per-track price of $0.50. Although young adults spend more in terms of dollars, this market segment only comprises 15% of total music sales.

Putting prices with market reach and revenue (Table 7), pricing singles at $0.99 confirms an optimal situation for OMPs. Though OMPs could price songs for less at say $0.50 and reach more consumers, revenue falls from $1,772.89 to $992.20. In terms of revenue from singles, the $0.99 price is optimal.

Traditionally, the record industry has made most sales from albums rather than singles. In 2003, for instance, CD album sales in the U.S. accounted for $11.2 billion compared to $35.9 million for CD singles. With most retail outlet offering albums from $12+, OMPs must price their

Table 5 Top Subscription Music Providers

	Catalog	Monthly Subscription	Format	Features				Usage Rights		
				D	S	FR	PR	L	B	P
Napster	700,000	$9.95	WMA, 128 kbps	✓	✓		✓	3	✓[a]	✓[b]
Musicmatch On Demand	700,000	$9.95	WMA, 128 kbps		✓	✓	✓	1		
RealRhapsody	600,000	$9.95	WMA, 128 kbps		✓	✓	✓	1	✓[a]	

Services: D = downloads; S = on-demand streaming; FR = free radio; PR = premium radio.
Usage Rights: L = number of simultaneously authorized computers for listening; B = tracks can be burned to CD; P = tracks can be transferred to portable devices

(a) Tracks to be burned to CD must be purchased for an additional fee. (b) Requires a Napster-To-Go subscription.

Source: Company web sites as of 24-Sept-2004.

Table 6   Consumer Willingness to Pay for a Music Download

	<$0.25	$0.26 to $0.50	$0.51 to $1.00	$1.01 to $2.00	>$2.00
Ages 18 to 24	17%	30%	34%	11%	7%
Total online adults	8%	19%	48%	12%	14%

Data from response to the survey question: Think about a song by your favorite band or artist. How much would you be willing to pay to download that song, which you could use as you wish? (Type in exact dollars and cents.)

Source: Jupiter Research/Ipsos-Insight Music Survey (7/03), n = 298 (users ages 18 to 24 willing to pay for song downloads), n = 1,870 (overall online adults) (US only)

Table 7   Price Expectations

Single Track

	$0.01	$0.50	$0.99	$1.49	$1.99
Percentage of market reached	89%	82%	74%	42%	39%
Potential revenue*	N/A	$992.20	$1,772.89	$1,514.44	$1,878.16

Album

	$0.01	$5.00	$9.00	$12.00	$16.00
Percentage of market reached	87%	72%	36%	12%	2%
Potential revenue*	N/A	$9,129.60	$8,216.64	$3,651.84	$811.52

* Potential revenue is used to illustrate how pricing impacts market size and affects profits, and is calculated by multiplying the percentage of market reached by the survey sample, and, in turn, by the unit price.
Base: 2,400+

Source: Jupiter Research (8/02)

albums competitively. Most OMPs, including the iTMS, price albums starting at $9.95. According to the survey data results, should OMPs lower album prices to $5.00, they would double the market reached and raise revenues. Reducing prices below this point captures a larger market share but at the expense of revenues.

At present, OMPs enjoy a 40 percent cut for every dollar earned on a download sale, or approximately double the traditional retail margin. This can partially explain why OMPs can provide steeper discounts—such as for multiple purchases—than their retail counterparts. Record companies also earn a larger share of revenues from downloads at 30 percent versus 12 percent from traditional retail. Surprisingly in both cases, the artist earns 12 percent of the revenues regardless of whether the sale was conducted online or at a retail outlet (Informa Media, 2003).

While record labels have been largely reluctant to embrace online music, digital sales are actually more beneficial for revenues and copyright management.

First, online music sales eliminate, or otherwise drastically reduce distribution costs since the purchased music is transferred electronically; this eliminates the need to actually produce a physical CD and ship the CD from the manufacturing plant to the retail outlet.

Second, OMPs allow better advertising through personalization. OMPs maintain a past record of songs purchased or streamed and can therefore provide better recommendations on other songs, artists, and genres that a user might like.

Lastly, selling music online permits better copyright monitoring. Contrary to the common notion that songs purchased from OMPs will lead to higher piracy rates, music purchased online is actually more secure since the songs have embedded Digital Rights Management (DRM) policies. These policies, though not completely piracy-proof, do offer greater than security than a CD acquired from a retail store, which can be easily ripped to a hard disk as unprotected MP3 files. Most of the songs offered by the leading OMPs are encoded using Microsoft's WMA file format which features DRM restrictions on burning, duplication, and dissemination. Other services like the iTMS use a proprietary DRM technology which restricts where and how the music can be played.

Conclusions

The continued popularity of Apple's iTMS suggests that so long as OMPs can continue to provide competitive pricing and relatively free DRM policies, there is a market for digital music.

A-la-carte downloading services will outpace subscriptions nearly 3-to-2 in 2008, most likely because users favor the ability to pick the tracks they want to purchase rather than be subject to a bundled good. Nearly all of the leading OMPs provide the same standard set of features and pricing: $0.99 singles, a dedicated software client, and an extensive catalog of music from the five major labels plus many independents. Differentiation and success in the digital market will depend on content exclusivity, community-oriented tools, file compatibility, and less restrictive usage rights.

Despite file sharing concerns, consumers have indicated that they would be willing to pay for music insofar as they can enjoy the same usage rights as with a purchase. An additional layer of differentiation for OMPs is thus universal playback compatibility and transfer to a portable media device.

Apple's iTMS has prospered

where its predecessors have failed largely due to its across-the-board 99-cent price and relatively free DRM policies. Having been able to secure uniform licenses with the major labels, Apple is able to provide a catalog-wide price as opposed to prices which may vary depending on the licensor. Moreover, Apple's usage policies give consumers many of the same freedoms they had enjoyed before with retail music.

As the music industry begins to better understand the market opportunities for pure digital sales, new entrants will emerge and competition for users and dollars will intensify.

References

"Broadband Worldwide 2004: Internet Usage Statistics and Growth Trends." *eMarketer* (2004).

Bernoff, Josh, et al. "The Digital Music Consumer: 2003." *Forrester Research* (2003).

Card, David. "99-Cent Digital Singles: Market Opportunity and Value Chain." *Jupiter Research* (2004).

Dyson, Simon. *Music on the Internet*. London: Informa Media Group, 2003.

Gartenberg, Michael and Peter Sargent. "The Apple Music Store: An Example of DRM Done Correctly." *Jupiter Research* (2003).

Haley, Colin C. and Robyn Greenspan. "Downloaders Will Pay for Ownership." *Clickz Network: Solutions for Marketers* (March 2003): 16 pars. 8 July 2004. <www.clickz.com/stats/markets/retailing/article.php/2110001>

Madden, Mary and Lee Rainie. "The Impact of Recording Industry Suits Against File Swappers." *Pew Internet & American Life Project* (2004): 6 pgs. 16 September 2003. <www.pewinternet.org/pdfs/PIP_File_Swapping_Memo_0104.pdf>

—. "The State of Music Downloading and File-Sharing Online." *Pew Internet & American Life Project* (2004): 13 pgs. 16 September 2003. <www.pewinternet.org/pdfs/PIP_Filesharing_April_04.pdf>

Oberholzer, Felix and Koleman Strumpf. "The Effect of File Sharing on Record Sales: An Empirical Analysis." March 2004: 51 pgs. Harvard Business School and UNC Chapel Hill. 10 July 2004. <www.unc.edu/~cigar/papers/FileSharing_March2004.pdf>

"Pricing Music Downloads: Online Music's Volume-Driven Future." *Jupiter Research* (2003): 11 pars. 16 September 2003. <www.jupiterresearch.com/bin/item.pl/research:concept/105/id=86125,pos=2/>

Recording Industry Association of America. "2003 Year-end Statistics." 20 June 2004. <www.riaa.com/news/newsletter/pdf/2003yearEnd.pdf>

Footnote

[1] See Oberholzer & Strumpf (2004) table 13. The effect of downloads grows more positive for more successful albums (higher quartile).
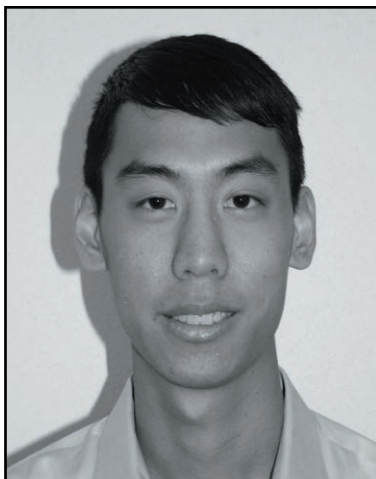
Albert Lin

Mentors: Professor Timothy Bresnahan, Dept. of Economics
Albert Lin is a junior majoring in Economics and co-terming in Management Science and Engineering. As a DJ and music producer, he is interested in understanding how technology will affect music creation, distribution, and live performance. He first worked with Professor Timothy Bresnahan on research involving web browser versioning and search engine technologies in the fall of 2003. Albert is grateful to Professor Bresnahan for his sponsorship, encouragement and professional review, to the URO for the Chappell-Lougee research grant, and to Jupiter Research and Informa Media for their invaluable research assistance.

Art and Literature

The Danger of the Abstract

Sameness and Isolation in Margritte's Golconda *and Koestler's* Darkness at Noon

Govind Persad

The middle of the 20th century featured two contemporaneous movements, Communism and Surrealism. Despite their radically different spheres of influence -- Communism in the political and Surrealism in the aesthetic -- the two movements shared core ideas, such as the devaluation of individual uniqueness. This paper attempts to connect both movements by considering a critique of each: Rene Magritte's post-Surrealist painting Golconda (1953), and Arthur Koestler's fictionalized depiction of a Communist prison in his short novel Darkness at Noon (1940).

I. Isolation and Separation

Rene Magritte's *Golconda*[1] draws the viewer into a group of increasingly identical men. This progression, from a real, concrete individual (the viewer) to *Golconda*'s undifferentiated mass of abstract mankind, is opposite to Peter Viereck's suggestion that Rubashov evolves from abstract mankind to concrete man in *Darkness at Noon*[2]. *Golconda* draws attention to the men in its center, only to propel the viewer's eyes outward along geometric lines to their fellows. As the viewer's visual focus progresses outward from the center, the exact physical distance between the men establishes their separation. David Sylvester states:

> This surprising separateness, in isolating them and denying their solidarity, seems to make each one vulnerable, for, whereas the familiar prospect of a body of impassive men uniformly dressed side by side in a solid phalanx is a paradigm of human solidarity, the unfamiliar prospect of a similar body of men widely distanced from one another becomes a paradigm of human isolation.[3]

Despite their sameness, these men are utterly disunited. Sylvester suggests that these men have not placed themselves in this separation voluntarily but "seem to have been moved, like chessmen."[4] This idea of men as chess pieces, to be sacrificed in order to achieve a larger goal, suggests the view, portrayed in *Darkness at Noon,* that humans are merely a means to a greater end.

Rubashov and the other prisoners, separated into isolation cells, embody Sylvester's idea of *Golconda* as representing separation. As the prisoners' human individuality, and hence their shared humanity, vanishes, so does their ability to communicate with one another. Rubashov's cell visually isolates him from his fellows. Though his jailers observe him through the judas, through it he can only see a miniscule slice of the hallway. His interactions with others, like the conversations he and his neighbors conduct by tapping on the wall, are solely auditory; the visual component of communication has been lost. Even the isolation cells are described through sounds alone: "In all the white-washed cells of this honeycomb in concrete, men were simultaneously arising from their bunks, cursing and groping about on the tiles, yet in the isolation cells one heard nothing."[5] This loss of visual contact appears even more strongly in Rubashov's medical condition. The pain emanating from the broken-off root of "the right eye-tooth which was connected to the eye-nerve orbitalis"[6] suggests the privation of Rubashov's vision. His eyes, made useless by the prison, become unusable.

The gazes of *Golconda*'s inhabitants suggest a similar visual isolation. No man in the painting looks at another. Each stares out of the painting to draw the viewer in, or off to the side to push the other men

Rene Magritte's *Golconda*.

away. This lack of visual connection between men repeats itself as the viewer's eyes traverse the painting's first layer, that consisting of the most foregrounded and therefore most detailed men. Held fast in the plane of the painting, its men can only look out of their world, not into it.

Resisting the foreground men's gazed invitations to join their number, the viewer's searching eyes next alight on the monolithic building that occupies the painting's lower half. The building's form is all straight lines and right angles, a neatly arranged and staggered row of rectangular windows themselves divided into smaller rectangles. Each gray, rectangular windowbox, devoid of even one irregular, beautiful flower, is a microcosm of the entire building; it imprisons all, within and without the painting, in its geometric regularity.

Like *Golconda*'s gray monolith, *Darkness at Noon*'s prison reduces its inhabitants' specific names to generalized numbers. This transformation—begins at the cell door:

> They reached cell no. 404. Above the spy-hole was a card with his name on it, Nicolas Salmanovitch Rubashov. They have prepared everything nicely, he thought; the sight of his name on the card made an uncanny impression on him.[7]

The card, however, does not denote Rubashov's individuality but rather strips it from him. His name, like the card, remains outside the cell. Within the cell, he is no. 404, and his neighbors are 402 and 406. Rubashov observes the mechanical monotony of the prisoners' walks outside, just as his own are restricted to a demarcated six and a half steps[8]. In the quadratic alphabet[9] of the prison, even words become numbers: "A familiar row of figures flashed into Rubashov's mind: 2-3, 1-1, 4-3, 1-5, 3-2, 2-4…'Hare-lip sends you his greetings.'"[10] The quadratic alphabet, with its letters confined in 5 rows of neatly numbered

cells, recalls both the windows of *Golconda* and the cells of the prisoners. The prison mathematically incarcerates all concepts, from language to individuality. Rubashov, as he strives to express the concept of "I," can only tap out the numbers "2-4."[11] This moment exemplifies the systematic replacement of the real world with an abstract model, an ideal which will be crucial in both works' attempt to provide an internal critique of each movement, one that appropriates the movement's values, such as abstraction, in order to show that these values fail by their own lights.

II. Abstraction and Sameness

The triumph of the logical, and hence of the abstract, suffuses *Darkness at Noon*. "We replaced vision by logical deduction," writes Rubashov, alluding to the visual bleakness of his prison as well as the creative poverty of Communism[12]. Sight becomes unnecessary to one who has foresight. Rubashov's old idea rests purely on this system of logical deduction, but logical deduction can only be assumed, not proven.[13] Logic works only because humanity must be simplified to the whole numbers of a mathematical model: as Rubashov queries, "Was such an operation justified? Obviously it was, if one spoke in the abstract of 'mankind'; but, applied to man in the singular, the cipher 2-4, the real human being of bone and flesh and blood and skin, the principle led to absurdity."[14] In the real world, logic, by assuming an abstract humanity, ignores human individuality.

Golconda similarly confines humans to an abstract ideal. Magritte said of his painting, "It may have been a result of a research into the

'question of space.' I don't believe a dream would go this far, and indeed the problem is an important one"[15] Magritte's solution to the problem of space in *Golconda* appears to be absolute geometrical perfection. Every window, every man, every shadow is evenly spaced. Sylvester suggests, "If you imagine what it would be like to have a world where everyone was neatly arranged with equal space, it would look pretty much like *Golconda*."[16] The quest for abstract perfection denies the variation of human experience. *Golconda* fits into Robert Hughes' generalization that Surrealism is "doomed to failure when it runs up against the real world."[17]

The abstract elements in both works cannot handle the concrete complexity of their human dimension, so abstraction works to remove that dimension. Viereck states of *Darkness at Noon*, "Were it only political history, *Darkness*…would remain narrow and two-dimensional. Its artistic vision is what adds a third dimension of human depth"[18] Communism wants political history without human depth. The artistic beauty of the fictional narrative, according to Viereck, gives *Darkness at Noon* a third, human, dimension. Yet, reducing the human dimension to the abstract, as *Darkness at Noon* often does, creates a work of art even while satisfying Communism's requirements. Humanity, once represented in two dimensions, can be spread out on the painter's canvas. Magritte himself recognizes that painting abstracts the human dimension: "In contrast to the image we have of them in real life when they are concrete, the painted image aroused a very strong feeling of an abstract existence."[19] Thus *Golconda*, too, fulfills its function as a narrative

of the abstract.

The canvas's reduction of humanity to two dimensions takes place in *Darkness at Noon* as well as in *Golconda*. Before Rubashov is imprisoned, the portrait of No. 1 haunts his mental world:

> the colour-print of No. 1, leader of the Party, which hung over his bed on the wall of his room—and on the walls of all the rooms next to, above, or under his; on all the walls of the house, the town, and of the enormous country for which he had fought and suffered.[20]

The repetition of No. 1's picture turns *Darkness at Noon* into a giant *Golconda*: the same houses with the same picture of the same man inside. The other painting depicted in *Darkness at Noon,* that of the first Communist Party congress, freezes Rubashov's own past in a two-dimensional frame[21]. This painting combines face, number, and name; it catalogues and depicts the past of Rubashov and his Communist contemporaries.

The eventual disappearance of this painting, deleting Rubashov from the past, presages Rubashov's death. Rubashov, in prison, cannot recall the faces of the dead; he forgets the faces of his own contemporaries, the "old guard" depicted in the painting who joined the Communist movement alongside him. Even those he killed become faceless. Rubashov, as he prepares to destroy Richard, finds that "Richard's features began to become hazy in the growing dusk."[22] Similarly, Bogrov's face grows increasingly indistinct as his death fades into the past.[23] By the end, Rubashov cannot even distinguish his own face from No. 1's; he has become faceless to himself through

his atrocities, just as his victims were faceless to him. Rubashov asks himself, "Yes, how many of yours have I had shot, I wonder?…there must have been something between seventy and a hundred."[24] When all men are the same, no one man has any value; when no one man has any value, no group can have value. Rubashov's crimes deface his victims both literally and metaphorically. His final crime, his false confession, effects the ultimate loss of face by equating him and No. 1 in his own mind.

The progression of faces in *Darkness at Noon*, from many to one to none, corresponds with the natural progression of *Golconda*, now moving upwards from the building to the sky. The greater separation of the largest men in *Golconda*, those who appear most prominent in the page, parallel Rubashov's memory at *Darkness at Noon'*s beginning, when men are still distinct from one another. As the viewer goes deeper to the next level of the painting, all the faces fade into similarity. Like the end of *Darkness at Noon*, the third, least foregrounded, layer of men in *Golconda* depicts facelessness. The men of *Golconda*'s third layer are almost entirely faceless, with their identical bowler hats their only distinguishing features.

In *Golconda* and Magritte's other paintings, the bowler-hatted man is a fusion of isolation and sameness: "A metaphysical loneliness, bordering on the spiritual and the stoical, surrounds the bowler-hatted man."[25] This isolation also enables the bowler-hatted man to represent all men. He provides an ideal background for our projected ideas:

> The bowler-hatted man is rather like Ulrich in Robert Musil's long novel The Man Without Qualities: he has

given up his qualities as a man might give up the world. Magritte's bowler-hatted man is more like a figure in a book than a human being, but a figure with all the inessential elements left out.[26]

In their abstraction, the bowler-hatted men of *Golconda* fulfill the ideal requirements of both Surrealism and Communism. They are everyone and no-one: "by putting the hat on, you are assured of being the perfect example of the average man."[27] This idea repeats in Hughes's assertion that *Golconda* parallels the experience of an average Belgian clerk on an average Belgian day[28]. Yet, as the exact average, the bowler-hatted man assures his separation from human uniqueness. Even in their conformity, whether large or small, faced or faceless, these men stand alone.

Finally, completely enveloped by the sameness of both works, the viewer searches for a way out, and finds one. The building on the right is pristine, entirely untouched by the cascade of men. At its base, Magritte has inscribed a subversive message: his own name. Magritte's signature, and the building above it, assert the individuality of the artist. This building goes against the other themes of *Golconda*, extending unbroken and untouched by man from the canvas's top to its bottom. Magritte's signature is not abstract geometry, but definite and original, referring to a particular individual: Magritte himself. His assertion of his unique name suggests his ability to transcend logic and abstraction. Magritte steps outside his canvas, as an actor might step out of the stage's mythical "fourth wall," to address his viewer personally. This act on Magritte's part shows that human connection and individual uniqueness can together change an image from one of isolation to one of friendship, and illustrates a way out of the box of abstraction. Meanwhile, the building's unending character, not vertically contained by the boundaries of the canvas, suggests another way to escape abstraction: by means of infinity. Infinity, a quantity not expressible in purely numerical terms, proves the key to resisting the grip of logic's faceless abstraction.

III. Infinity and Individuality

Another of Magritte's paintings, his later (1963) *La Reconnaissance Infinie*[29], is both a companion and antagonist to Golconda. *Reconnaissance*'s triple meaning (translatable into English as "gratitude," "recognition," or "reconnaissance"), shows that the connection between the two floating bowler-hatted men depicted in the painting involves mutual appreciation and seeing the other as a person, as well as creating the possibility of infinite understanding or sight. Through this later work, which utilizes the same tropes as *Golconda*, Magritte suggests an infinite connection that cannot be enumerated or quantified, but only observed. As such, *La Reconnaissance Infinie* provides a further critique of the abstract.

The power of the infinite also undermines the abstract in *Darkness at Noon*: "The infinite was a politically suspect quantity, the 'I' a suspect quality. The Party did not recognize its existence. The definition of the individual was: a multitude of one million divided by one million."[30] Rubashov regrets his ignorance of the infinite, feeling that it establishes his guilt. This concept of infinity appears in the French title of *Darkness at Noon* as well, "Le Zero et l'Infini"[31], which effectively fuses the individual and infinity. Infinity, inexpressible in numerals, challenges the numeric system of Darkness at Noon. The Party can enumerate a billion people, but infinity cannot be reached by counting.

Rubashov's victims, like Magritte, resist the abstract through their names. Unlike the faceless prisoners, Arlova, Richard, and Little Loewy assert their identities in Rubashov's mind, refusing to bow to the arithmetic reduction of prison: "The vision of Arlova's legs in their high-heeled shoes trailing along the corridor upset the mathematical equilibrium."[32] These names, like Magritte's in *Golconda*, cannot be reduced to numeric equations. Finally, as Rubashov prepares to kill Bogrov, Bogrov's naming of Rubashov himself awakens Rubashov to existence as an individual rather than as part of a mass, just as Magritte's signature in Golconda personalizes the painting and reaches out to the viewer.

The end of *Darkness at Noon* returns to the infinite. "A second, smashing blow hit him on the ear. Then all became quiet. There was the sea again with its sounds. A wave slowly lifted him up. It came from afar and travelled sedately on, a shrug of eternity."[33] The strike to his ear separates Rubashov from his final sense, hearing, while the sea suggests submersion in infinity. As *Golconda*'s viewer disappears into the infinite blue behind the last bowler hat, Rubashov slips into an eternal ocean. Yet even death does not absolve Rubashov of moral responsibility. Rather, it suggests a question Koestler poses in his autobiography:

> I was, of course, in prison and might be shot. But this was immediately succeeded by… floating on my back in a river of peace, under bridges of

silence….The I had ceased to exist….If everybody was an island, how could the world be a concern of his?[34]

No human begins as an island. Abstraction's danger stems from its balkanization of humanity, dividing humans into units and adding them back together to form an archipelago of identical, yet isolated people. Magritte may in fact use his identical men "to challenge the conformity he pretends to illustrate."[35] However, depicting the abstract contains its own seductive danger, that the artist or observer may accept the depiction without recognizing its irony. Embracing human uniqueness and connection, as Rubashov does at *Darkness at Noon*'s end when he rejects the isolated vision of Communism can avoid the isolation and sameness of *Golconda*'s universe. Magritte's recognition that "an object never has the same function as its name or its image,"[36] an admission of abstract art's limitation, is similarly important in fighting the conflation of world, word, and image depicted in *Darkness at Noon*. Only when they subvert the abstract ideals they both illustrate do *Golconda* and *Darkness at Noon* become agents for positive change rather than depictions of abject abstraction.

This aspect reflects the iconoclastic nature of each work, not only with respect to the abstract ideal but also with respect to the movements they critique. Both Communism and Surrealism, centering as they do on the manifestoes of Marx (1848/1890) and Breton (1924) respectively, demand at least a modicum of consistency between the form and content of their members' works and the ideals set down in the manifesto. By utilizing the literary and artistic tropes of abstraction espoused by their respective movements to critique those movements, *Golconda* and *Darkness at Noon* upset this idea of consistency and provide a powerful internal challenge, rather than merely an external objection, to Surrealism and Communism.

Works Cited

Gablik S. <u>Magritte.</u> Greenwich, CT: New York Graphic Society, 1970.

Hughes R, writ. and narr. <u>The Threshold of Liberty.</u> Part 4 of <u>Shock of the New</u>. Videotape. New York: Time-Life Video, 1980. 52 min.

Koestler A. <u>Darkness at Noon.</u> New York: Bantam, 1968.

Magritte R. <u>Golconda.</u> 1953. The Menil Collection, Houston.

Magritte R. <u>La Reconnaissance Infinie</u>. 1963. Leslie and David Rogath Collection, Greenwich, CT. .

<u>The Mystery of Magritte.</u> CD-ROM. Brussels: Virtuo, 1996.

Ollinger-Zinque G and Leen F, eds. <u>Magritte, 1898-1967.</u> Trans. Tiffany Barnard, Ted Alkins Davidson, and Margaret Clarke. Ghent: Ludion, 1998.

Soby JT. <u>Rene Magritte.</u> New York: Museum of Modern Art/Doubleday, 1965.

Sylvester D. <u>Magritte: The Silence of the World.</u> New York: Abrams, 1992.

Torczyner H. <u>Magritte: Ideas and Images.</u> Trans. Richard Miller. New York: Abrams, 1977.

Viereck P. Introduction. <u>Darkness at Noon.</u> By Arthur Koestler. New York: New American Library, 1961.

Endnotes

[1] Magritte R. <u>Golconda.</u> The Menil Collection, Houston, 1953.

[2] Viereck P. Introduction to <u>Darkness at Noon</u> by Arthur Koestler. New York: New American Library, 1961: xiv.

[3] Sylvester D. <u>Magritte: The Silence of the World.</u> New York: Abrams, 1992: 296.

[4] Ibid., p. 296.

[5] Koestler A. <u>Darkness at Noon.</u> New York: Bantam, 1968: 9.

[6] Ibid., p. 46.

[7] Ibid., p. 8.

[8] Ibid., p. 12.

[9] Ibid., p. 19.

[10] Ibid., p. 158.

11 Ibid., p. 205.

12 Ibid., pp. 80-81.

13 Ibid., p. 81.

14 Ibid., p. 206.

15 Torczyner H. Magritte: Ideas and Images. Trans. Richard Miller. New York: Abrams, 1977: 200.

16 Sylvester, Magritte, 296.

17 Hughes, Robert, The Threshold of Liberty.

18 Viereck, Introduction, Darkness at Noon, vii.

19 Ollier-Zinque G and Leen F, eds. Magritte, 1898-1967. Ghent: Ludion, 1998: 45.

20 Koestler, Darkness, 4.

21 Ibid., p. 47.

22 Ibid., p. 34.

23 Ibid., p. 117.

24 Ibid., p. 44.

25 Gablik S. Magritte. Greenwich, CT: New York Graphic Society, 1970: 156.

26 Ibid., 156.

27 The Mystery of Magritte, CD-ROM. Brussels: Virtuo, 1996.

28 Hughes, The Threshold of Liberty.

29 Magritte R. La Reconnaissance Infinie. Leslie and David Rogath Collection, Greenwich, CT, 1963.

30 Koestler, Darkness, 208.

31 In English: "Zero and Infinity."

32 Ibid., p. 116.

33 Ibid., p. 216.

34 Viereck, Introduction, Darkness at Noon, xiv.

35 Soby JT. Rene Magritte, New York: Museum of Modern Art/Doubleday, 1965.

36 Ollier-Zinque and Leen, Magritte, 1898-1967, 47.

Acknowledgements

Govind Persad

The most challenging part of writing this paper was definitely the 40-minute walk with a backpack full of oversized art books from the Palo Alto library back to the Marguerite stop on University Avenue. Try it sometime! I am now a junior double majoring in philosophy and biology and still taking literature courses, with my favorite being Joshua Landy's Proust seminar. When I return from Stanford-in-Oxford, I hope to write an honors thesis on biomedical ethics -- then it's probably off to graduate school in philosophy, unless I decide that I enjoy sleeping under a roof and regular meals more than I love Kant and Descartes. Outside of class, I enjoy reliving the first-year experience (minus the actual writing) vicariously via my SLE tutees. I also help edit Stanford's undergrad philosophy journal, *The Dualist*, and, pre-Oxford, frequently played late-night games of pool at Phi Sig.

Mapping the Protestant Experience on Human Flesh Imagery in Jean de Léry's Historie d'un voyage fait en la terre du Brésil

How Léry examines nudity and cannibalism in the context of the French Reformation

Liz Rountree

In 1550, France celebrated its new relationship with the South Atlantic by importing fifty Tupinamba natives to "simulate their battles" and entertain King Henry II and Catherine de Medicis on the eve of the colonial era (Denis 3-4). It was "a truly curious monument for the two countries" (19) because what began as a festival symbolizing international prowess and fascination with the exotic cannibal ritual became an omen for France's own religious "battles" culminating in the Saint-Barthelemy's Day Massacre.

Jean de Léry, a member of the Huguenot resistance, made this comparison by using his experience in Brazil to condemn the violence of the French Reformation. In the process of writing *Histoire d'un voyage fait en la terre du Brésil*, Léry synthesizes the Renaissance method of scientific observation and humanism. Léry subverts his own colonial perspective and sees the humanity of the Tupinamba instead of the contemporary trend to portray them as mythical half-human cannibals waiting to be civilized.

He uses their humanity to criticize his own French culture and experience of religious turmoil. Paolo Carile describes Léry's ability to "apply this reversal of perspectives which cultivates a respect for others over the self so as to establish a critical distance, a space that can become ironic, by relation of the one's own self and one's own culture" (Carile 31). Léry integrates his observations of Tupinamba culture with questions of why and how they compare to his own culture. This culminates in his juxtaposition of cannibalism with the Saint-Barthelemy's Day Massacre (Jurt 62). Cannibalism as a metaphor for the massacre ironically reflects the schism of the Reformation over transubstantiation and consubstantiation (Léry Voyage 68).

Léry fashions a new ethnographic method through his vocabulary and thematic organization. He also condemns the violence of the French Reformation by comparing his Brazilian and French experiences. From the conception of colonial endeavors to the final published edition, all events and characters involved show the intimacy between the production of this text with the Protestant experience in sixteenth century France. I compare Léry's text with that of his contemporary, André Thevet [Singularities of the French Antarctic]. I have focused on their discussion of nudity and cannibalism because both concern "the human flesh" and symbolize the crisis of the Eucharist and opposing interpretations of flesh.

A quick briefing on Léry and Thevet and their involvement with the colonial process

Gaspar de Coligny, a Protestant working for the Catholic Crown, funded the French endeavor in Brazil

to establish a refuge for persecuted Protestants on the Island of Dieppe (just off the Brazilian mainland) (Larcarde 82). The colonists united under a banner of tolerance, and the party included twenty-two year old Jean de Léry. The veil of tolerance soon disintegrated with controversy over transubstantiation resulting in the starvation of the Protestants by Nicolas Durand Villegagnon, a Catholic and vice admiral to Coligny (Jurt 55). The Protestants fled to the mainland, where Léry encountered the Tupinamba and observe their culture and in a sense flee his national identity (Léry 55). He was no longer confined to a French identity and thus the French definition of humanity.

Léry compares "the inhumanity that [Villegagnon] uses against these people" to the "treason he used against us, sending us from our outpost to the land of the savages" (85). As Frank Lestringant writes: "There he was rejected and annihilated, now on the exterior, like the Amerindians" (Lestringant Studi 82): as a Protestant, Léry was worth no more than a savage to his people. His transition from Frenchman to savage, this dehumanization, influences Léry to rethink the definition of humanity.

Reading Léry's new ethnography

Léry and Thevet reveal their authorial intentions when they approach nudity. Thevet's description comes from his colonial quest to civilize (Jurt 58). He emphasizes the inseparability of nudity from bestial anti-Christian behavior:

> without law, religion, without any civility, but living like irrational beasts, … always nude, men as much as women, and until they come to be converted by Christians will, little by little, shed their brutality, to dress civilly and

more humanely (Thevet 134-5).

Thevet asserts that a colonial effort will "clothe" this naked, bestial society ("without law, religion, or any civility") with a "fashion," or society fit for humans.[1]

Léry, conversely, uses nudity to show that the Tupinamba are actually quite fit humans. He describes their nudity as natural and their bodies as well-formed (Léry 96). He emphasizes their humanity: "Understand, imagine, a naked man, well-proportioned and well formed in his parts." Asking his readers to "imagine … a man" immediately asks them to imagine a human, as they know one, a European, as opposed to imagining an "irrational beast." Combined with his attention to detail he constructs a realistic portrait of the Tupinamba people. Léry eventually uses this portrait to make a social critique of his own people.

Léry writes an equally realistic account of cannibalism. Neither romanticizing nor condemning it, he situates "these murders" (221) in the context of Tupinamba warfare ceremony. Léry observes that the Tupinamba ate their prisoners of war "more for vengeance than for taste" (220). Thevet casually describes how they ate the Portuguese: "for without any reasons, the Savages of the country killed [the Portuguese] and ate them as they do to their enemies" (Thevet 111). Before his audience learns about "the Savages," he affirms the prevalent mythology. Thevet confuses his discourse on cannibalism with other examples of monstrosity around the world, interchanging the word cannibal with savage, black, idolatrous, and beasts. It is difficult to understand which people are the Tupinamba. His inability to render a clearer portrait of who practices cannibalism or how it is practiced merely emphasizes the

cruelty of cannibalism, propagating the mythology of somewhere, dark(er) skinned half-humans roasting themselves.

Léry eases his readership into the Tupinamba world. He cushions abstract concepts such as religion and cannibalism between chapters on social culture and language, which allows his European audience to understand the Tupinamba society as equally complex, comparable to Europeans, and more human, and vice versa. This attention to presentation and ability of observation make this text what Joseph Jurt calls the "*chef d'œvre de la litérature ethnographique*" (Jurt 53). This new method of ethnography gives him an epistemological framework against which he can contrast his own French culture and the Reformation.

Reading Léry's critique of the Reformation

Léry could have written an anthropological account of the Tupinamba in his Histoire d'un voyage in 1558, upon the return of his voyage. The first publication twenty years later (1578), post-dating the failure of the French colony at Dieppe, Thevet's two publications *Singularitez* (1558) and *Cosmographie Universelle*[2] (1575), and Léry's own *Histoire memorable de la ville de Sancerre which recorded the famine of Sancerre* and Saint-Barthelemy's Day Massacre (1572). Léry incorporates these events into *Histoire d'un voyage* and in doing so transcends ethnography and shows the inhumanity of his own culture and the turmoil of the Reformation.

Léry's *Histoire d'un voyage fait en la terre du Brésil*, presumably about his voyage, has a remarkable similarity to *Histoire memorable de la ville de Sancerre* (Larcade 51). Léry invokes his first publication, in which he spares no details of the grotesque

suffering in the fourth religious war of 1572-73 (Larcarde 52), with the title of the second.[3] Géralde Nakam describes how Léry incorporates Histoire memorable as way to foil the violence of the Reformation against the Tupinamba culture:

> The voyage to Brazil, the crisis of Sancerre, represent both exile and solidarity. Since, because of rekindled polemics, the witness can no longer be anonymous: he incarnates, with his own eyes, the exoticism of the noble Brazilian and the Sancerrian tragedy (Nakam, 130).

In *Histoire d'un voyage*, Léry's criticism of the French Reformation begins with a study of nudity and culminates in a comparison of cannibalism to the Saint-Barthelemy's Day Massacre.

Where Thevet sees nudity as a physical and mental lack in the presence of abundant material, Léry asks why the French wear clothes. He describes Tupinamba nudity as free of voyeurism because there is no "dishonesty to look at nude women" (Léry 114). In contrast, he argues the elaborate clothing of French women is sexually excessive (Jurt 61) and is "without comparison, the cause of much misconduct than the ordinary nudity of female savages" (Léry 114). Léry concludes that instead of wearing clothing "for honesty and necessity," the French wear clothes "for glory and worldliness" (Léry 115). The nudity of the Tupinamba is not immodest because clothing hardly signifies modesty in his own culture. This criticism of French clothing seems mundane, but it introduces a radical new "reversal of perspectives" (Carile 31). Léry uses the Tupinamba culture to question the norms of his own culture (as opposed to the traditional "normative" style

of ethnography like Thevet who evaluates the foreign to his French standards (Jurt 56)). Léry establishes this "reversal" early in his text so when he arrives at more abstract concepts like cannibalism, his readers understand, or at least are acquainted with, his method of comparison.

On cannibalism, Thevet again compares "their culture" to "French culture." Always instilling a mortal fear of cannibals who "have no difficulty in killing a Christian and eating him]," Thevet compares how cannibals eat in French terms: they eat "like we here eat beef and mutton" (Thevet 140). Comparing cannibalism to French cuisine happens before any discussion on cannibalism or how it fits into its greater context of war ritual both isolates and belittles the cannibal practice to a choice of diet, rendering it morally simple, easy to condemn. Léry intentionally safeguards against this kind of thinking by contextualizing cannibalism in the Tupinamba culture of war and into the greater context of universal human cruelty.

Léry does this by describing Tupinamba cannibalism with the violence of the French Reformation. He shows that barbarity is universal human experience by emphasizing the extent of Protestant suffering and Catholic hypocrisy. He concludes his chapter on cannibalism by introducing this concept of universal cruelty:

> To those who will read these horrible things, exercised ritually between the barbarous nations of the land of Brazil, can come somewhat closer to understanding such dangers among us ourselves (Léry 228).

Léry argues that learning about these cruel practices can make "those who will read" think about violence "among" their own culture.

He continues with an example of cruelty within French civilization by referencing the famine of Sancerre: "it would be best to cut the throat all in one stroke, then to do it so as to let it languish" (228). In essence, Protestant suffering was so great that it seemed painless in comparison. He implies that the Tupinamba are in some respects (morally) more human in killing their enemies rather than making them suffer. He also points out the hypocrisy of Catholic condemnation of cannibalism by describing Transubstantiation: "they eat, nevertheless, not grossly, but instead spiritually, to eat the flesh of Jesus Christ" (68). In this metaphor the Catholics become cannibals.

Léry evokes the horror of Saint-Barthelemy's Day Massacre to develop this relationship between Transubstantiation and cannibalism. He introduces the massacre:

> And without going far from which France? (I fled from France) during the bloody tragedy that began at Paris on the 24 of August, 1527 (229).

He implicitly asks why condemn violence abroad when such cruelty exists at home. He develops his metaphor of the Catholics as cannibals, "why is no horror expressed for those for whom the hearts and other parts of the body that were eaten by their furious murderers?" (229) Léry argues that the Catholics were not satisfied eradicating the Protestants but had to make the Protestants suffer by "eating" their faith and their hearts. Léry substitutes the verb "to eat" for murder to postulate why killing can be condoned in a society that condemns cannibalism. He ironically demonstrates the ultimate equivalency between killing someone and eating someone: the death of the victim.

However, he defines this

"butchery of the French people" as "more barbarous and cruel than that of the savages" (229). because it concerns French people killing French people (as opposed to Tupinambas killing their prisoners of war). To him, what matters is the pain, suffering and murder of "their parents, neighbors and compatriots" (230), because Saint-Barthelemy's Day Massacre represents the possibility of localized cruelty and violence. It also sets a precedent, crucial to the time of the developing nation-state, for lines to be drawn within French borders declaring who is and who is not a real Christian, who is and is not part of the nation.[4] Léry argues there is no universal definition of Christianity just as there is no universal definition of humanity.

Nudity and cannibalism: lessons in tolerance

In *Histoire d'un voyage*, Léry writes of a world where, naked or not, one is either a cannibal or the victim of cannibalism. Whether one is a nude Tupinamba native, a suffering Protestant or a Catholic supporter of the Reformation in sixteenth century France, cannibalism exists both literally and metaphorically. Léry is able to make this connection because he defines cannibalism in his ethnography of the Tupinamba and then compares it to how the Protestants suffered at the hands of the Catholics. In his long discourse on nudity and cannibalism he attempts to negate the mythology and render the Tupinamba people human. Rendering this image allows his European audience to expand their concept of humanity. He uses *Histoire d'un voyage* as his looking glass, to reflect the inhumanity of his own society. Léry's new anthropology, based on a broader concept of humanity, also

reflects a shift in religious tolerance that his Protestantism represents, and vice versa (Lestringant, "La Renaissance"). The evolution of anthropology and Christianity in the sixteenth century can simultaneously be characterized by a new impetus for tolerance.

Léry's new "tolerant" ethnography based on a more liberal understanding of humanity is perhaps an effect of his expulsion from the Catholic colonial project, making him feel more affection and affiliation for the Tupinamba. But his text resonates with the greater philosophical tensions of the Renaissance. Jean Céard describes this new mode of liberal thinking in the politico-religious climate of the Renaissance: "This tension between the honest quest for unity and the surprising face of the admirable variety of the world" (Céard 77).

The sixteenth century has both this "obstinate quest of unity," symbolized by the use of religion in the process of nation building, and this "admirable variety of the world," symbolized by the explosion of European discovery. This new and sudden diversity negated one of the fundamental building blocks holding up European society, namely the Christian story of Genesis. Tension between pre-New World and post-New World concepts of human origin called for a reevaluation of the fundamentals holding society together. Lestringant names this discourse "hierarchical thought" against "the dissipated and the disparate:" where the former subordinates "diversity to unity" featuring "a center, origin, and unique source," and the latter is "stuck in the concrete" and "it never goes beyond its own generalizations" (Lestringant "La Renaissance" 75-6). Thevet can be seen as an example of the former because his descriptions of nudity

and cannibalism always include a comparison back to French Catholic values, his operational center, his one truth. Léry, on the other-hand, questions French Catholic values in comparison to the Tupinamba, simultaneously validating Tupinamba as part of humanity and debunking the concept of a universal definition of humanity.

Léry professes a similar concept in his *Advertissement de l'Autheur* in the final publication of *Histoire d'un voyage* (1611). As Géralde Nakam introduces it, "where the conquistador says 'more to have than to see'" (Nakam Léry, 44), Léry writes in his last edition, "MORE TO SEE THAN TO HAVE" (Léry *voyage* 398). The colonizer would rather have everything before he would see or understand what is in his grasp. It does not matter what it is, as long as it is subordinated to the colonizer, his religion or his one "unity." Léry, on the other hand, says that it is better to know than to have: that it is better to understand the Tupinamba culture then to claim to it, that it is better to accept Protestantism then to eradicate it.

Léry writes his Protestant experience on a canvas of human flesh by using his description of nudity and cannibalism in Brazil as a comparison to the cruelty of his own French culture. His new ethnography comes from a mélange of his own experience as an "other" (being part of a persecuted religious minority): the civility he felt amongst the Tupinamba when expelled from the French colonial project, as compared to the barbarity of his own people during the Reformation. Because of his own experiences he advocates a new religious tolerance that broadens the definition of Christianity. His situational ability to dissociate from his own culture allows him to operate

outside its traditional epistemological framework. His vision, unrestricted by the national colonial agenda, allows him to expand the concept of humanity and thus create a new ethnography founded on principles of tolerance. Léry's *Histoire d'un voyage* embodies the transitions in Renaissance thought: away from the sub-human cannibal to the human, away from the heretic to the accepted Christian, and away from universal truth to concepts of diversity.

Works Cited

Adonias, Isa. Mapa: Images of Brazil's Territorial Genesis. Rio de Janeiro: Odebrecht, 1993.

Augeron, Mickaïl. "Coligny et les Espagnoles à travers la course." Coligny, les Protestants et la Mer. Eds. Acerra, Martine; Martinière, Guy. Paris: Presses de l'Université de Paris-Sorbonne, 1997 (155-176).

Carile, Paolo. "Jean de Léry, les Protestants et l'Écriture sous l'Ancienne Régime." Jean de Léry, "Histoire d'un voyage fait en la terre de Brésil" : actes de la journée   d'étude sur Jean de Léry, 6 novembre 1999. Ed. Gisèle Mathieu-Castellani . Paris : Equipe Tradition antique et modernités" de l'UFR "Science des textes et documents", (1999). 31-42.

Castelnau Letoile. "Le voyageur et le missionaire: analyse de deux regards sur les Indiens du Brésil." Histoire d'un voyage en la terre du Brésil, Jean de Léry: Journées d'étude (10 et 11 décembre 1999). Ed. F. Argod-Dutard. Bordeaux: Centre Montaigne, 2000, 35-48.

Céard, Jean. La Nature et les prodiges: L'insolite, en France, au XVIe siècl. Genève: Droz, 1977, p. 229-408.

Denis, Ferdinand. Une Fête Brésilienne célébrée à Rouen en 1550. Paris: J. Techener, Libraire, 1850.

Gilmont, Jean-Françcois. Jean Crespin, un éditeur réformé du XVIe siècle. Genève: Librairie Droz, 1981.

Jurt, Joseph. "Le Chef-d'oeuvre de la literature ethnographique, Le Voyage faict en la terre du Brésil..." Dispositio/n. Ann Arbor, Mich., Dept. of Romance Languages, University of Michigan. 17 (1992): 53-70.

Léry, Jean de. Histoire de Sancerre. Au Lendemain de la Saint-Barthélemy Guerre Civile et Famine. Paris:Éditions Anthropy, 1975 (171-361).

_____. Histoire d'un voyage fait en la terre du Bresil. Ed. Jean Claude Morisot. Geneve : Droz, 1975.

Lestringant, Frank. André Thevet. Genève: Librairie Droz, 1991.

_____, and Marie-Christine Gomez-Géraud. D'encre de Brésil: Jean de Léry, écrivain. Orléans: Paradigme, 1999.

_____. Écrire le monde à la Renaissance : quinze études sur Rabelais, Postel, Bodin et la littérature géographique. Caen : Paradigme, 1993.

_____. "La Renaissance, nous et les autres: Note sur l'anthropologie du XVIième siècle." Studi Francesi, 2003 Jan-Apr; 47 (1 (139)): 71-83.

Nakam, Géraldo. Au Lendemain de la Saint-Barthélemy Guerre Civile et Famine. Paris:Éditions Anthropy, 1975.

The Holy Bible: The New King James Version. Atlanta: Thomas Nelson Publishers, 1992.

Thevet, André. Les Singularitez de la France Atarctique. Ed. Paul Gaffarel. Paris, Maisonneuve & Cie, 1878.

_____. Portraits from the Age of Exploration. Ed. Roger Schlesinger, trans. Edward Benson. Chicago: University of Illinois Press, 1993.

Vergé-Franceschi, Michel. "L'amirauté de France dans la 2ième moitié du XVIe siècle: un enjeu entre Catholiques et Protestants." Coligny, les Protestants et la Mer. Eds. Acerra, Martine; Martinière, Guy. Paris: Presses de l'Université de Paris-Sorbonne, 1997 (35-43).

Footnotes

[1] To his credit, Thevet later describes a more innocent nudity, "[Ils] viuent touts nuds ainsi qu'ils sortent du ventre de la mere" (Thevet 141), but this initial aggressive attack on Tupinamba nudity and how it relates to their lack of culture defeats the purpose of attempting to make himself appear the "benevolent viewer."

[2]   Jean-Claude Morisot suggests that Léry's publication in 1578 is a direct response to Thevet's Cosmographie

Universelle. In this publication Thevet blamed the failure of French colonial endeavors in Brésil on the Protestants (Morisot XI). Also included in this publication were conversations with natives in Canada, where Thevet supposedly never landed (Schlesinger 5).

[3] The last line of his introductory sonet in Histoire memorable summarizes his thesis of universal humanity based on a universal experience of tragedy: "Tout l'univers de hideux en partage" (Léry Sancerre 177).

[4] Frank Lestringant points out the irony of a nation that condemns cannibalism yet will "<<vomit>> l'intrus au lieu de l'integerer" because this practice of expulsion from within is often attributed to "des societes anthropophages" (Lestringant "La Renaissance" 82).

Liz Rountree

Liz Rountree, class of '06, is majoring in Comparative Literature major with French and Spanish. She is about to embark on her honors thesis which will continue her study on human imagery in the "New World" in sixteenth century texts. She is also a sailor from San Francisco and currently captain of the varsity Stanford sailing team.

Intellectual Property Law

Patents and Pharmaceutical Drugs
The Need for Change

Ahmad Siddiqi

The tension between two goals – encouraging medical research while keeping medical treatments affordable – is as yet unresolved. This paper first examines how implementation of the WTO's TRIPS agreement initially restricted health access in developing countries and assesses whether these concerns have been subsequently addressed. It then examines the effects of patents on stimulating medical research, and discusses cases of market inefficiencies and failure. It goes on to consider several approaches that attempt to fix some of these shortcomings, and concludes by suggesting six points essential to a comprehensive strategy in tackling these issues.

Arguably the most controversial aspect of the World Trade Organization's Agreement on Trade-related Aspects of International Property Rights (TRIPS) is over the issue of patents for pharmaceutical drugs. To their proponents, patent rights are essential to encourage innovation, as the virtual monopoly they create allows firms to extract greater profits from drugs that they invent. Opponents, however, point out that patents result in higher prices, making essential medicines less affordable. In this paper, I address the issues of health access and medical research in turn. I first examine how implementation of the TRIPS agreement restricted health access in developing countries and assess, in light of the outcomes at Doha in July 2001 and later on August 30th, 2003, whether these concerns have been sufficiently addressed. I then turn to the second issue, examining the effects of patents on stimulating medical research and discuss cases of market inefficiencies and failure. I go on to discuss several approaches

that attempt to fix some of these shortcomings. Finally, I suggest six points essential for a successful strategy in tackling these issues.

Barriers to Distribution

Background: The TRIPS agreement

The TRIPS agreement was brought into effect on January 1, 1995. The agreement unambiguously extended the scope of intellectual property rights, declaring in Article 27 that "patents shall be available for any inventions, whether products or processes, in all fields of technology."[1] Moreover, patent rights would be "enjoyable without discrimination as to the place of invention, the field of technology and whether products are imported or locally produced."[2] These provisions were in part a response to the lack of protection for pharmaceutical products in countries such as India, whose 1970 Act of Parliament granted "patent rights only to manufacturing processes, rather than to the end products themselves,"

allowing Indian firms to 'reverse-engineer' the production process and manufacture generic copies of drugs.[3] Pharmaceutical companies in the US and EU had campaigned in favor of the agreement, as the provisions ensured that any company that invented a drug would have exclusive rights over its production and distribution for the duration of the patent.[4] However, by preventing cheap generic copies of a drug under patent from being produced, the TRIPS agreement had serious ramifications for public health. The ability of patents to restrict access to treatments for diseases became glaringly evident with the rapid spread of HIV/AIDS.

The AIDS pandemic

The first cases of Acquired Immune Deficiency Syndrome (AIDS) were documented in 1981 in the US and UK.[5] By 1997, UNAIDS reported that 30 million people were infected with HIV.[6] There were estimated to be 3 million cases of HIV/AIDS in India alone.[7] In Sub-Saharan Africa,

the effects were far worse; already, the region accounted for more than half the HIV/AIDS cases worldwide.[8] That ratio has scarcely improved since.[9]

The development of antiretroviral (ARV) treatments for HIV/AIDS appeared "to be a remarkable success story, with more than 20 highly effective medicines developed in the past 17 years."[10] However, the average yearly cost of antiretroviral treatment using patented drugs is $10,000-15,000 per person.[11] Of the population of developing countries, roughly one billion people live under a dollar a day.[12] Given that per capita health expenditures in low income developing countries average $23 per year, only a fraction of those infected with AIDS in Third World countries could afford such treatment.[13]

South Africa authorizes generics

Responding to the alarming growth of HIV infections in the country, South Africa passed a law in 1997 giving the government "blanket powers to…produce or import cheap alternatives to the brand-name drugs for HIV and other diseases."[14] Since producers of generics do not need to invest money in research, they can sell at a fraction of the cost of patented drugs. As Appendix A shows, this fraction can be as little as six percent.

Thirty-nine pharmaceutical companies raised a court challenge to prevent South Africa from implementing the law. John Barton, former chair of the UK Commission on Intellectual Property Rights, observes that this move "became a public relations debacle for the industry."[15] A multitude of NGOs, including Oxfam and Medecins Sans Frontieres (MSF),

sharply criticized the pharmaceutical companies for attempting to restrict health access. With pressure mounting, the companies eventually dropped their challenge, "after threats that the amount of public support for the development of the relevant drugs would be publicized in the hearings."[16] Despite the victory, the South African government has been criticized for moving too slowly on resolving its AIDS epidemic since then.[17]

Brazil's fight against AIDS

In 1996, the Brazilian government began offering free ARV therapy to people with AIDS. As the costs of this program grew, the government expanded its health budget and increased its production and import of generics. Brazil also used the threat of compulsory licensing – authorizing companies to produce generic copies of patented drugs – to force patent-holders to cut prices.[18] As a result, drug prices in Brazil are much lower than other countries and the government has succeeded in cutting AIDS mortality rates by 50 percent.[19]

In response to Brazil's actions, the US filed a complaint with the WTO in early 2001, accusing the government of violating TRIPS. As in the case of South Africa, the move backfired. In April 2001, all members of the UN Human Rights Commission except the US supported a Brazilian resolution asking nations not to "deny or limit equal access for all persons to preventive, curative or palliative pharmaceutical or medical technologies used to treat pandemics such as HIV/AIDS."[20] The US withdrew the complaint in June 2001, saying it would pursue the issue in bilateral talks.

Doha and the August decision

The spread of the AIDS pandemic, along with highly publicized cases such as Brazil and South Africa, contributed to increased awareness on an international level that patents threatened to severely restrict access to health in developing countries. This awareness contributed to the outcome at Doha in November 2001, hailed as the development round of the WTO. In response to the issue of patents on pharmaceutical drugs, the conference released the "Declaration on the TRIPS agreement and public health" on November 14. The declaration did not offer any substantial revisions to TRIPS; rather, it recognized flexibilities that already exist in the agreement, such as Article 31, which establishes the procedures by which a compulsory license may be granted. According to 31(f), a compulsory license must be "predominantly for the supply of the domestic market of the Member authorizing such use."[21] The Doha declaration stressed that TRIPS "should be interpreted and implemented in a manner supportive of WTO members' right to protect public health and, in particular, to promote access to medicines for all."[22] Moreover, "each member has the right to grant compulsory licenses and the freedom to determine the grounds upon which such licenses are granted."[23] Recognizing that due to Article 31(f), countries lacking sufficient drug manufacturing capacity would have difficulty employing the compulsory licensing clause, the agreement referred this issue to the Council for TRIPS to resolve by the end of 2002. The decision, eventually made on August 30, 2003, allowed all least-developed countries (LDCs), as well as any other member that provided prior notification, to import generics from abroad by issuing a

compulsory license. The requirements of 31(f) would be waived for this purpose.[24]

An imperfect solution

With the conclusion of these agreements, the world is moving towards a consensus that differential pricing – making the costs of medicine cheaper for people in developing countries than in developed countries – is a solution to the issues of health access raised by patents. The argument makes economic sense, as developing world markets provide minimal incentive for research; "the total market of the poorest countries…is on the order of 1 percent of the global pharmaceutical market."[25] Moreover, the logic of patents is that they allow prices to be artificially high on the understanding that increased profits will be fuelled back into research. As John Barton observes, "It seems reasonable that the burden of these costs, which benefit all of humanity, should fall more heavily on the wealthy than on the poor."[26] Nonetheless, while the Doha agreements were an important step towards promoting health access, several issues remain.

The pharmaceutical industry is correct to point out that patents are not the only obstacle to health access. There is a range of causes including poor health infrastructure, lack of sanitation, and shortage of funding for even generic drugs. Any comprehensive approach to tackling the problem will need to address all of these factors.

Moreover, despite the Doha agreements, patents still have the potential to restrict health access. A key concern is whether the production of generics will decrease as a number of developing countries fall under the patents regime on January 1, 2005. The Commission on Intellectual Property Rights notes:

"There will be no incentive, as now, for manufacturers in these countries to reverse engineer newly patented drugs and take the other steps necessary for manufacture and sale (including obtaining regulatory approval), because the domestic market would be closed. Thus the ready supply of generic substitutes for patented drugs now available will gradually disappear. Potential compulsory licensees would therefore have to charge a price closer to full economic cost (including start-up and manufacturing costs) as compared to the possibility of providing off-the-shelf generics at prices where start-up costs have already been amortised to some extent on the domestic market. Moreover, if the necessary investment is only triggered by the availability of a compulsory licence, there will inevitably be long delays before the drug actually reaches the intended patients."[27]

Addressing the economic question is as important as addressing the legal one. MSF has been campaigning for an "Article 30 exception" that would allow countries like India that manufacture generics to continue exporting to countries that need them.[28] When member countries of the WTO meet in Hong Kong in 2005, they would do well to ensure that any amendment to TRIPS, incorporating the August 30th agreement, also addresses concerns that a supply of generics be readily available for developing countries.

Incentives to Innovate

The logic behind patents is a form of social contract between society and firms. Society agrees to grant patent-holders a temporary monopoly, thus allowing them to set prices at an artificially high level, on the understanding that doing so will provide a strong incentive for innovation. In the case of an industry constantly creating new products, such as the pharmaceutical industry, the assumption is that much of the increased profits will be fuelled back into research, resulting in new treatments for diseases being discovered.

There is a general consensus that patents on pharmaceutical drugs provide a very strong economic incentive for research. The R&D requirements of creating new drugs are significant, and the potential payoff that a patent grants encourages companies to make the necessary investments. However, while one must acknowledge these incentives, it is essential that an optimal balance be struck between the benefits to society from increased innovation and the costs of paying the patent derived price. There are certainly reasons to question whether such a balance currently exists.

Inefficiency and waste

Pharmaceutical companies argue that a new drug may cost $500-$800 million to develop, and that perhaps only 30 percent of these drugs may make a good return, though as the Commission on IPR observes, these figures are "contentious."[29] Still, there is no doubt that the industry is compensated well, having occupied

3

the top of the Fortune 500 list for most of the last two decades.[30] And yet it is far from ideal in terms of developing new treatments for diseases. According to MSF, the industry spends twice as much on marketing as on R&D and "It is in any case a vastly wasteful mechanism for encouraging innovation, since only one tenth of drug sale profits is plugged back into R&D."[31] Moreover, as a study in *The Lancet* showed, "68 percent of all new chemical entities marketed worldwide in the last 25 years were 'me-too' products, representing little or no therapeutic gain."[32] As such, substantial resources may be diverted from meaningful innovation, without any significant drop in prices.

An ounce of cure is worth a pound of prevention

In addition to inefficiency, the market system can also be a source of research bias. According to the International AIDS Vaccine Initiative (2002), private industry HIV/AIDS vaccine R&D expenditure totaled a meager $50 – $70 million.[33] By contrast, even conservative estimates put the cost of developing a single antiretroviral drug at $200 million. In a working paper for the Center of International Development (CID) at Harvard University, Michael Kremer and Christopher Snyder explain the gap in expenditure by presenting an economic model that predicts "drug revenue could exceed vaccine revenue by more than a factor of four" in the case of sexually transmitted diseases.[34] This market bias carries disturbing implications for the prospect of developing vaccines for such diseases.

Neglected diseases: A clear case of market failure

However efficient research may be, patents cannot provide an incentive for R&D into treatments for diseases that do not have a market. The lack of a market does not imply a lack of customers; it implies an inability to pay. Treatments exist for AIDS, as pharmaceutical companies could be certain of recovering R&D expenditures in the developed market, regardless of whether poor nations purchased their drugs. There is no such guarantee for tuberculosis, pediatric AIDS, malaria and other tropical diseases that affect people primarily in poor nations. The Commission on IPR observed:
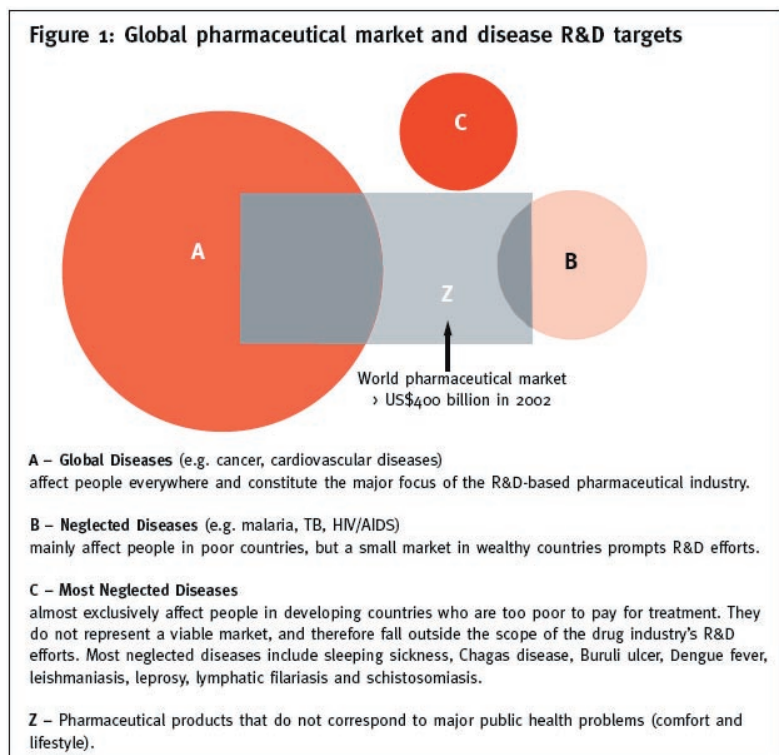
> "So what role does IP protection play in stimulating R&D on diseases prevalent in developing countries? All the evidence we have examined suggests that it hardly plays any role at all, except for those diseases where there is a large market in the developed world…The heart of the problem is the lack of market demand sufficient to induce the private sector to commit resources to R&D."[35]

The Commission estimates that less than five percent of pharmaceutical R&D worldwide is spent on diseases that predominantly affect the developing world.[36] The imbalance is shown in Figure 1.

Possible Solutions

Awareness of these shortcomings has contributed to the emergence of a number of alternative approaches that attempt to correct the shortcomings in the research system. Many of these initiatives however, came up only recently and have not yet been evaluated for effectiveness. Others remain under-funded; the pharmaceutical industry still funds most global R&D for health, estimated in total at $75 billion in 2002. A few of the most interesting initiatives

Figure 1: Global pharmaceutical market and disease R&D targets

World pharmaceutical market > US$400 billion in 2002

A – Global Diseases (e.g. cancer, cardiovascular diseases) affect people everywhere and constitute the major focus of the R&D-based pharmaceutical industry.

B – Neglected Diseases (e.g. malaria, TB, HIV/AIDS) mainly affect people in poor countries, but a small market in wealthy countries prompts R&D efforts.

C – Most Neglected Diseases almost exclusively affect people in developing countries who are too poor to pay for treatment. They do not represent a viable market, and therefore fall outside the scope of the drug industry's R&D efforts. Most neglected diseases include sleeping sickness, Chagas disease, Buruli ulcer, Dengue fever, leishmaniasis, leprosy, lymphatic filariasis and schistosomiasis.

Z – Pharmaceutical products that do not correspond to major public health problems (comfort and lifestyle).

Adapted from '*Fatal Imbalance: The Crisis in Research and Development for Drugs for Neglected Diseases*', Médecins Sans Frontières/DND working group, 2001

follow:

The Global Fund to Fight AIDS, Tuberculosis and Malaria was created on the initiative of UN Secretary General Kofi Annan, governments of African countries and the G8, in 2001. The Fund is a partnership between governments, civil society, the private sector and affected communities. Its purpose is to "dramatically increase resources to fight three of the world's most devastating diseases."[37] French President Jacques Chirac has called for a commitment of $3 billion a year to the Fund so that it is able to provide sufficient resources to the various initiatives it supports.[38]

Public-private partnerships (PPPs) and not-for-profit drug development initiatives bolster research on neglected diseases by "matching existing capacity, expertise and resources in both the public and private sector on specific projects or diseases."[39] The Bill and Melinda Gates Foundation is a major contributor to PPPs focusing on neglected diseases.[40]

In addition, many innovative approaches have been suggested but not yet initiated. Notable among these is the idea of establishing a vaccine purchase fund, as suggested by Jeffrey Sachs, noted economist and former Director of the CID at Harvard. In "The Case for a Vaccine Purchase Fund," Sachs, et al. argue that while there exists significant evidence that development of vaccines for AIDS, tuberculosis and malaria is scientifically feasible,[41] there is no economic incentive for pharmaceutical companies to invest in their development. Thus, the authors propose establishing an artificial market: governments contribute to the fund according to their means, but no actual expenditure is made until some firm develops a reliable vaccine. The fund would then purchase the vaccines to immunize children in low-income countries where these diseases are prevalent. The authors argue that there are two advantages to this approach that do not exist in public funding of research: there is a stronger incentive for researchers to focus on vaccine development, and the cost of failures is borne by firms rather than taxpayers.[42] Sachs and his colleagues estimate that prices of $10 – $40 should provide sufficient incentive, resulting, if vaccines for all three diseases are discovered, in immunization of 160 million children per annum at a yearly cost of $1.6 – $6.4 billion to the fund.[43] This they see as a modest cost for eradicating diseases that together account for nine percent of all deaths.

Conclusion: The Way Forward

Clearly, in the complex matters of health access and medical research, one cannot expect any single approach to work across all countries and all diseases. But turning ideas into results will require coordination and direction. At present the World Health Organization is the one entity capable of fostering the growth of such a global commitment. The WHO is in a unique position due to its status as the only organization legally mandated as responsible for global health issues. It has significant achievements to its credit in the past, such as the eradication of smallpox, efforts to improve sanitation in developing countries, and establishment of biological standards for food and drugs. Recently, the WHO helped coordinate an "unprecedented" international effort to contain the Severe Acute Respiratory Syndrome (SARS) outbreak in 2003.[46] Thus, the WHO should take on a leadership role, coordinating global health efforts with long-term vision, and a concrete but flexible plan for implementation. Details of such a plan are best left to the organization, but its success will rely on adherence to the following six principles.

First, the WHO should set an R&D agenda for the world, prioritizing diseases based on which ones constitute major global health concerns. This global research agenda should be dynamic, with the WHO periodically reassessing the burden imposed by diseases and adjusting priorities as necessary. Once the agenda is set, governments should promote research into diseases deemed high priority.[47]

Second, the WHO should campaign for differential pricing between developed and developing countries to be established as routine in the case of essential drugs. Differential pricing could be arranged for by pharmaceutical companies or by the availability of generics, though developing countries should have the flexibility of both options. The organization should also lobby for an amendment to TRIPS in line with the waiver of Article 31(f), and it should stress the need to ensure a sufficient supply of generics, by amendment or other means.

Third, the WHO's regional offices should continue to help developing countries work on infrastructure, availability of treatments, sanitation and other potential barriers to health access. It should pressure governments of these countries to increase health spending and match these efforts with commitments from developed countries to provide funds, transfer of technology, or skills and expertise. The WHO should also initiate programs

to facilitate the establishment of pharmaceutical manufacturing sectors in LDCs and other countries that lack such capability.

Fourth, the WHO should continually assess the efficacy of various initiatives by governments, NGOs, PPPs and other groups that aim to address some shortcoming in health care or the current system of research. The WHO can use its existing Civil Society Initiative (CSI), set up in June 2001, for this purpose. CSI teams should assess the effects of such initiatives on a yearly basis, bringing to public attention the best practices for experience sharing, and as projects for donor funding. CSI should also recommend approaches that are especially effective and that the WHO could expand to a wider scope of problems or regions than originally envisioned.

Fifth, the WHO will have to continually review the impact of patents on global health. The World Health Assembly established the Commission on Intellectual Property, Innovation and Public Health in 2003. The Commission should be made into a permanent body, capable of carrying out its mandate through required financial and legislative support. Patents are not an inalienable right, but a means towards striking the best balance between innovation and cost. To determine that balance, countries must be able to assess the efficiency of the system. Thus, it would be in the interest of developed nations to assist the Commission with its investigations.

Sixth, the WHO should promote technology transfer and sharing of research outcomes in order to 'fast-track' innovation. Where revealing research outcomes may conflict with R&D incentives, the WHO should ask the Commission to determine the ideal balance.

In conclusion, successfully combating disease requires a willingness to put significant resources into efforts. In this age of globalization we must recognize a global responsibility. The HIV/AIDS epidemic, the SARS outbreak; these serve as poignant reminders of the need for concerted action. No single country is capable of eradicating diseases that know no national boundaries, and no single country can hope to remain insulated indefinitely from the world's problems. As we see major shortcomings in the market-oriented system of medical research, we should increasingly move towards the recognition that global disease cure and prevention is, in simple economic terms, a public good. Only through a sustained and substantial public-sector effort to complement the current privately funded research can a realistic solution be found.

Appendix A: Tables and Data
Section A1: Comparison between cost of patented drugs and Brazilian generics

ANTIRETROVIRAL DRUG PRICES
Difference between proprietary company offers and generic producer prices
   Price of AZT/3TC:
- GlaxoSmithKline (proprietary company), special discount price: US$ 2 per day
- FarManguinhos (generic): US$0.96 per day (52% cheaper)
   Price of Nevirapine:
- Boehringer Ingelheim (proprietary company): US$1.19 per day (3)
- FarManguinhos (generic): US$0.59 per day (50% cheaper)
   Price of AZT:
- GlaxoSmithKline (proprietary company): US$1.6 per day
- FarManguinhos (generic): US$0.09 per day (94% cheaper)
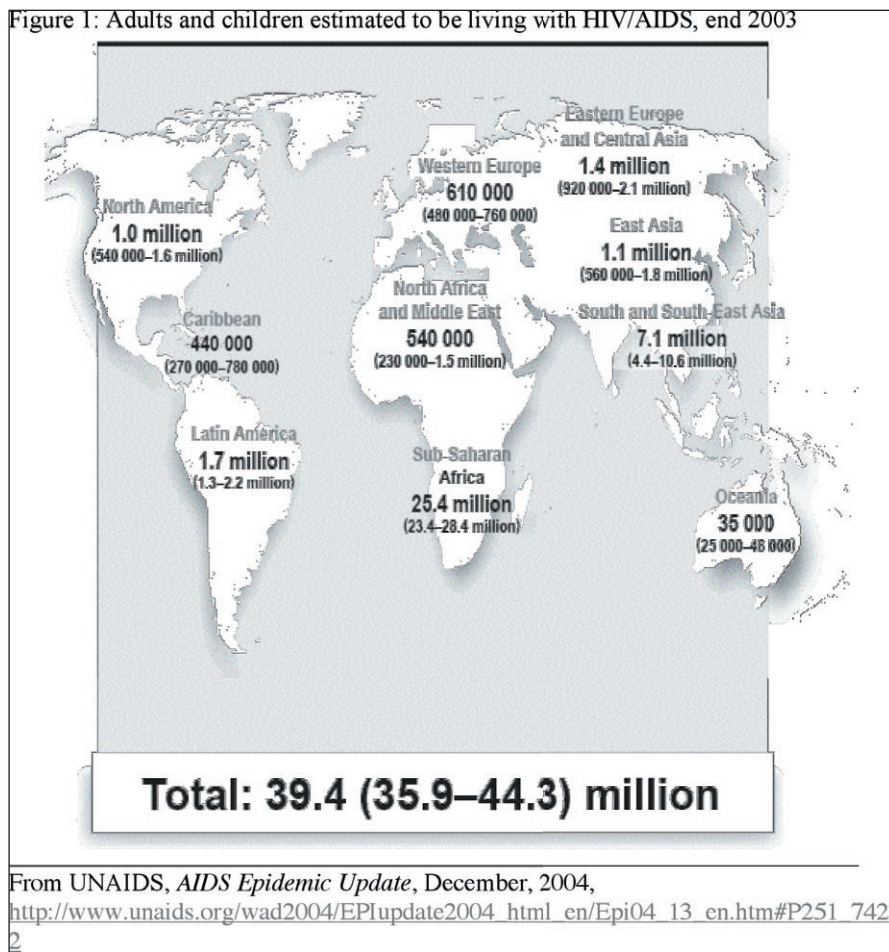   **Price of 3TC:**
- GlaxoSmithKline (proprietary company): US$0.64 per day
- FarManguinhos (generic): US$0.41 per day (36% cheaper)

From Medecins Sans Frontieres, "Brazilian Generic ARV Drugs in South Africa – the Backgound," 29th November, 2002

Section A2: Regional statistics for HIV/AIDS end of 2004

Figure 1: Adults and children estimated to be living with HIV/AIDS, end 2003

Eastern Europe and Central Asia
1.4 million
(920 000–2.1 million)

Western Europe
610 000
(480 000–760 000)

North America
1.0 million
(540 000–1.6 million)

East Asia
1.1 million
(560 000–1.8 million)

North Africa and Middle East
540 000
(230 000–1.5 million)

South and South-East Asia
7.1 million
(4.4–10.6 million)

Caribbean
440 000
(270 000–780 000)

Latin America
1.7 million
(1.3–2.2 million)

Sub-Saharan Africa
25.4 million
(23.4–28.4 million)

Oceania
35 000
(25 000–48 000)

Total: 39.4 (35.9–44.3) million

From UNAIDS, *AIDS Epidemic Update*, December, 2004, http://www.unaids.org/wad2004/EPIupdate2004_html_en/Epi04_13_en.htm#P251_74232

Endnotes

[1] World Trade Organization, Uruguay Round, 15th April, 1994, Annex 1C (TRIPS), Part II, Section 5, Article 27, Para 1, http://www.wto.org/english/docs_e/legal_e/27-trips_04c_e.htm#5

[2] Ibid.

[3] Newsweek International, "India Kicks the Habit," 22nd November, 2004, http://msnbc.msn.com/id/6478695/site/newsweek/

[4] Typically a period of twenty years.

[5] AVERT, "The history of AIDS up to 1986," http://www.avert.org/his81_86.htm

[6] AVERT, "The history of AIDS, 1993-1997," http://www.avert.org/his93_97.htm. HIV (Human Immunodeficiency Virus) is the virus that causes AIDS. 2.3 million people died of AIDS in that year alone.

[7] Ibid.

[8] Ibid.

[9] In 2004, UNAIDS estimates that 39.4 million are living with HIV/AIDS, out of which 60% (25.4 million) live in Sub-Saharan Africa, a region that accounts for only 10% of world population. UNAIDS, "AIDS Epidemic Update: December 2004 – Sub-Saharan Africa," http://www.unaids.org/wad2004/EPI_1204_pdf_en/Chapter3_subsaharan_africa_en.pdf

[10] Medicins Sans Frontieres, "MSF Campaign for Access to Essential Medicines: Briefing Note," November 2004, http://www.accessmed-msf.org/documents/MexicoR&Dbriefing.pdf

[11] Ibid.

[12] The World Bank Group, "World Bank: Partnerships in development," http://www.worldbank.org/progress/reducing_poverty.html

[13] Commission on Intellectual Property Rights, Final report, Chapter 2 (Health), November, 2002, http://www.iprcommission.org/papers/text/final_report/chapter2htmfinal.htm

[14] allAfrica.com, "Drugs' Giants Drop Case Against South Africa," 19th April, 2001, http://allafrica.com/stories/200104190053.html

[15] John H. Barton, "TRIPS and the Global Pharmaceutical Market," *Health Affairs*, Vol. 23, Issue 3, pp. 146-154.

[16] Ibid.

[17] AIDS drugs have been more expensive in South Africa than in Brazil, and the government took until April 2004 to make free antiretroviral therapy available to its citizens.
IRIN, UN Office for the Coordination of Humanitarian Affairs, "South Africa: Slow start of ARV rollout," 1st April, 2004, http://www.irinnews.org/report.asp?ReportID=40383&SelectRegion=Southern_Africa&SelectCountry=SOUTH_AFRICA (Note IRIN's views do not necessarily reflect those of the UN or its offices.)

[18] Supra n. 17.

[19] Appendix C shows how much cheaper several antiretroviral drugs are in Brazil. Also see Brazilian Ministry of Health AIDS Drugs Policy, http://www.aids.gov.br/assistencia/aids_drugs_policy.htm

[20] PBS, Online NewsHour, "Brazil: A Model Response to AIDS?" http://www.pbs.org/newshour/health/global/generics_wto.html

[21] Supra n. 1, Article 31(f).

[22] World Trade Organization, "Declaration on the TRIPS agreement and public health," 14th November, 2001, http://www.wto.org/english/thewto_e/minist_e/min01_e/mindecl_trips_e.htm

[23] Ibid.

[24] The resolution was delayed due to demands by the US, and initially the EU, that any exceptions be limited to a predefined set of diseases. These demands were eventually dropped on the condition that the Chair of the WTO General Council read out a parallel statement indicating that the agreement would only be used in good faith, for the purposes for which it was intended. Supra n. 15.

[25] Supra n. 15.

[26] Ibid.

[27] Supra n. 13.

[28] Medecins Sans Frontieres, "MSF calls upon governments to act upon recommendations of the Commission for Intellectual Property Rights," 12th September, 2002, http://www.accessmed-msf.org/prod/publications.asp?scntid=1792002143316&contenttype=PARA&%20
According to Article 30 of the TRIPS agreement, "Members may provide limited exceptions to the exclusive rights conferred by a patent, provided that such exceptions do not unreasonably conflict with a normal exploitation of the patent and do not unreasonably prejudice the legitimate interests of the patent owner, taking account of the legitimate interests of third parties."

[29] Supra n. 13.

[30] Supra n. 10.

[31] Ibid.

[32] Supra n. 10.

[33] Michael Kremer and Christopher M. Snyder, "Why is There No AIDS Vaccine?" CID Working Paper No. 111, October 2004, p. 1.

[34] Ibid, p. 3.

[35] Supra n. 13.

[36] Ibid. Or, to put it another way, only 13-16 of 1393 drugs developed between 1975-1999 were aimed at tropical diseases, which account for 10% of the global disease burden.

[37] The Global Fund to Fight AIDS, Tuberculosis and Malaria, http://www.theglobalfund.org/en/

[38] Ibid, http://www.theglobalfund.org/en/in_action/events/quotes/

[39] Supra n. 10.

[40] Examples of PPPs include the Medicines for Malaria Venture (MMV), the Global Alliance for TB Drug Development (TB Alliance), the International AIDS Vaccine Initiative (IAVI), the Institute for One World Health (IOWH), and the Global Alliance for Vaccines and Immunization (GAVI). The Drugs for Neglected Diseases Initiative (DNDi) is a not-for-profit initiative created in 2003 to carry out R&D into new drugs for neglected diseases, such as sleeping sickness, Chagas disease and leishmaniasis.

[41] Jeffrey Sachs, Michael Kremer and Amar Hamoudi, Center for International Development at Harvard University, "The Case for a Vaccine Purchase Fund," pp. 3-4. http://www2.cid.harvard.edu/cidmalaria/case_vpf.pdf

[42] Since firms are used to taking such risks when they invest in R&D, they would have ample incentive to develop vaccines, provided that the artificial market is large enough.

[43] Supra n. 41, p. 1.

[44] Supra n. 10.

[45] Ibid.

[46] Supra n. 10.

[47] Governments could do so by increased direct R&D funding, providing funds to firms or other organizations, or guaranteeing a market for research output. Governments could also consider imposing disincentives, such as increased taxes, on firms that pour resources into areas of little therapeutic worth.

Ahmad Siddiqi

Ahmad M. Siddiqi is a junior at Stanford University. He completed his high school education at Aitchison College in Lahore, Pakistan, and is currently pursuing a double major in International Relations and Mathematical and Computational Science. His interest in this research topic dates from an International Relations course that he took in Fall Quarter, entitled "Science, Technology, and Development," and taught by Professor Gili Drori. His non-academic interests include public-speaking, dancing, cricket and reading, of all sorts.

Marine Biology

Thermal Tolerance Limits of Heart Function in Tegula Snail Congeners of the Marine Intertidal

Emily Stenseng

Organisms inhabiting the marine intertidal zone face extreme environmental stresses. One such stress is temperature, and a species' thermal tolerance is determined by a suite of physiological factors. In this study, the thermal limits of heart function were determined for three marine snail species of the genus Tegula. T. funebralis is found in the low- to mid-intertidal, T. brunnea in the subtidal to low-intertidal, and T. montereyi in the subtidal zone. Using impedance electrodes, changes in heart rate were monitored in field-acclimatized and lab-acclimated specimens of each species, in response to thermal stresses in water. Two sets of critical temperatures were measured—Arrehnius Break Temperature (ABT, the temperature at which heart rate began a sharp decline), and Flat-line Temperature (FLT, the temperature at which the heart function ceased), and significant interspecific differences were observed for both. When subjected to heating, heart failure occurred at significantly warmer temperatures in T. funebralis, the highest-dwelling species. In response to cold stress, T. funebralis maintained cardiac function at lower temperatures than its subtidal congeners. For all species, ABTs of specimens lab-acclimated to 22oC were higher than those of 14oC lab-acclimated snails. T. funebralis, however, showed the smallest ABT difference between acclimation temperatures, which suggests a limited ability to acclimate. Although T. funebralis has a wider thermal tolerance range than its two lower-occurring congeners (related species), it regularly faces field body temperatures above its ABT, indicating that T. funebralis faces a larger threat from heat stress. Thus T. funebralis populations, and perhaps warm-adapted intertidal species in general, may be at highest risk in the event of global climate change.

The marine intertidal is one of the harshest environments on earth. Organisms inhabiting the transition from land to sea are subject to a host of environmental stresses, including temperature variation, dessication stress, wave force, and salinity [1, 2]. The daily tidal cycle is largely responsible for the drastic, rapid temperature fluctuations typical of the intertidal, exposing organisms to alternating terrestrial and aquatic conditions.

Many species inhabit discrete vertical zones in the marine intertidal [3]. These zonation patterns are established by abiotic factors, such as temperature and aerial exposure, as well as biotic factors like competition and predation [3, 4, 5]. Studies have shown that species in the mid- to high-intertidal zones experience more extreme environmental stresses than subtidal species. For instance, some high-dwelling gastropods can undergo body temperature increases of 20-25oC during low tide emersions [6], whereas many subtidal animals never face temperatures above that of the ambient seawater.

These drastic differences in thermal habitats exist only because intertidal organisms have differentially evolved adaptations for thermal tolerance. Numerous studies have shown that the upper vertical limits of species distributions are correlated with

thermal tolerance limits [7, 8, 9]. A suite of physiological, biochemical, and morphological factors contribute to the establishment of these limits-- mitochondrial respiration, action potential generation in nerve cells, heart function, protein stability, and heat-shock protein response are some such traits [10, 11].

One of these thermal tolerance parameters, heart function, is the focus of this study, and marine snails of the genus *Tegula* are the test subjects. *Tegula* is an ideal genus for heat stress research because of the abundance of closely related species that occupy diverse thermal habitats and discrete vertical zones [12, 13]. Three differentially distributed temperate *Tegula* congeners were selected (Fig. 1). *T. funebralis* inhabits the low- to mid-intertidal, and can encounter temperatures over 33°C when emersed (exposed to air). *T. brunnea* is found in the low-intertidal to subtidal zone. It is exposed to air only during low low tides and experiences maximum habitat temperatures of approximately 23°C. Finally, *T. montereyi* is an exclusively

subtidal species, which never exceeds ambient seawater temperatures.

The *Tegula* congeners selected for this experiment have been studied extensively with respect to thermal tolerance. Whole animal thermal limits have been determined—the LT50 values (the temperature at which 50% mortality is observed) for *T. funebralis*, *T. brunnea*, and *T. montereyi* are 42.5°C, 36°C and 36°C, respectively [14]. Interspecific differences in certain enzyme stabilities have also been characterized in these congeners. For example, the thermal stabilities of malate dehydrogenases from five *Tegula* congeners increase with increasing vertical position [10]. Results of *Tegula* heat shock protein (hsp) experiments correspond with these findings [14, 15, 16]. The synthesis of hsps in *T. funebralis* is initiated, maximized, and terminated at higher temperatures than in *T. brunnea* and *T. montereyi* [14].

These previous *Tegula* studies demonstrate clear correlations between maximum habitat temperatures and thermal tolerance limits. Ac-

cording to these prior findings, it is apparent that *T. funebralis*, the higher-occurring intertidal species, has many physiological and biochemical adaptations allowing it to endure warmer temperatures than its subtidal relatives. The purpose of this investigation was to elucidate the role of heart function among these thermal adaptations, by determining and comparing the upper and lower cardiac thermal tolerance limits of *Tegula* congeners from different tidal heights.

Materials and Methods

Collection and care of specimens

Tegula specimens were collected from the intertidal zone at Hopkins Marine Station in Pacific Grove, California (36° 36'N, 121° 54'W). All specimens were adults of medium to large body size (20-25 mm basal diameter), found in the upper region of their species' vertical distribution. Specimens designated as "field-acclimatized" were held in re-circulating aquaria containing ambient seawater (13-14°C) for 24 hours prior to experimentation.

For the laboratory acclimation experiment, snails were held in aquaria with water temperatures of 14°C and 22°C (± 0.2°C). Fifteen snails of each species were acclimated for 15-19 days and fed fresh kelp every three days. The water level of the tanks was kept high enough so that the animals could never emerse themselves. Keeping the specimens submerged at all times was important, because only the water temperature, not the air temperature, was experimentally controlled.

Heart Rate Measurements

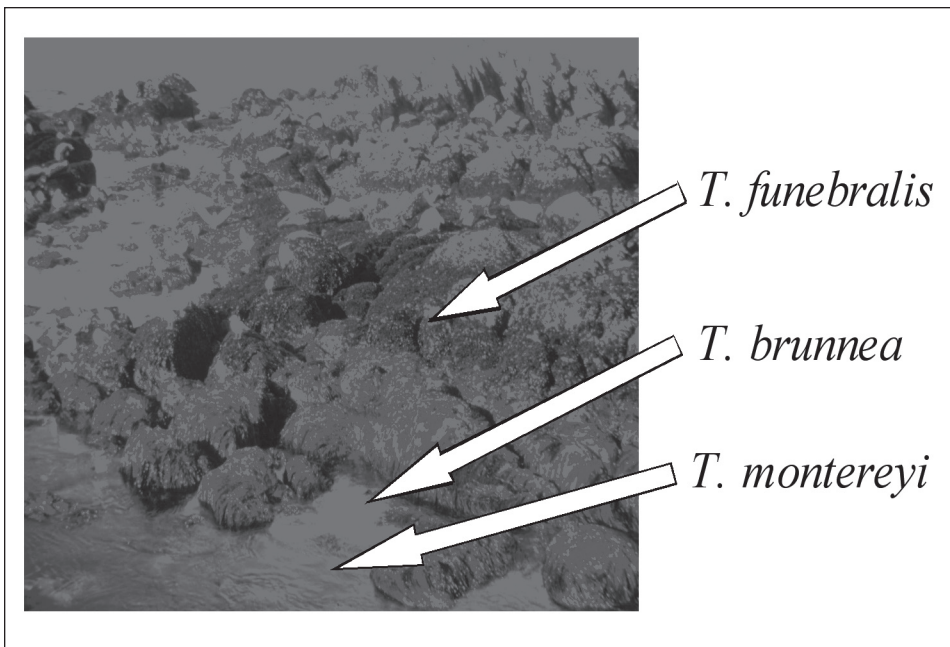Using a small hand drill, two holes were drilled into the shell of each snail, adjacent to the pericardial

Figure 1: The intertidal at Hopkin's Marine Station at low tide. Approximate vertical positions are shown for each of the species studied.

T. funebralis

T. brunnea

T. montereyi

space. Ceramic-coated copper electrodes were inserted into these holes and secured in place with SuperGlue™, positioned as close to the heart as possible. To prevent the snails from emerging during experimentation, the outer lip of each shell was glued to a clean glass microscope slide. The snails were suspended using metal clamps that held cork pieces glued to the top of each shell. The two electrodes in each animal were connected to an impedance pneumograph (a device that measures changes in electrical impedance in the tissue between electrodes), and the resulting signals were amplified and recorded using PowerLab™ Data Acquistion System (Fig. 2). Heart rate measurements were performed in filtered, aerated seawater, temperature-controlled by a LAUDA waterbath system. This setup accommodated a maximum of 6 animals per run.

For heat stress experiments, the waterbath temperature was held constant at 13°C for 1 hour, then increased by 1oC every 15 minutes, a rate that approximates natural daily temperature increases in the intertidal. After reaching 40°C, or once all specimens' hearts had failed, the water temperature was decreased as rapidly as possible, to assess recovery of cardiac function. Heart rate measurements were made every 7 minutes for the duration of the run.

Cold stress experiments were also performed, since emersed intertidal organisms can face extremely low air temperatures during winter months. Animals were held at 13°C for one hour, and then water temperature was decreased to nearly 0°C at a rate of 1°C every 10 minutes. As in the heat stress runs, heart rates were measured every 7 minutes. After heart failure, temperature was returned to ambient levels. Following every run,

the snails were weighed and placed in a re-circulating water bath at 13-15°C for 72 hours. Each animal was examined every 24 hours following experimentation to determine recovery and mortality rates.

In addition to temperature ramps, positive control runs were performed, to ensure that the conditions of the experiment did not cause deleterious stress contributing to the heart failure and mortality observed. In these runs, *T. funebralis* and *T. brunnea* heart rates were monitored for 7 hours in the absence of heat stress by holding the temperature constant at 13°C. In the week following the experiment, the control snails were closely monitored to assess recovery.

Data Analysis

Heart rates were expressed as beats per minute and plotted over time for every run, to visually assess cardiac responses to temperature change.

Arrhenius plots were generated for each animal by converting heart rate values to the natural logarithm of beats per minute and plotting these values against a temperature transformation (1000/K). Arrhenius break temperatures (ABTs) were determined using regression analyses to generate the best fitting line on both sides of an inflection point on the Arrhenius plot. The intersection of these two regression lines was taken as the ABT. Flatline temperatures (FLTs) were also determined as the temperature at which the last the measurable heart beat occurred. Critical temperatures, both ABTs and FLTs, were averaged for each species and acclimation condition, and statistical significance of comparisons was assessed using analysis of variance (ANOVA).

Results

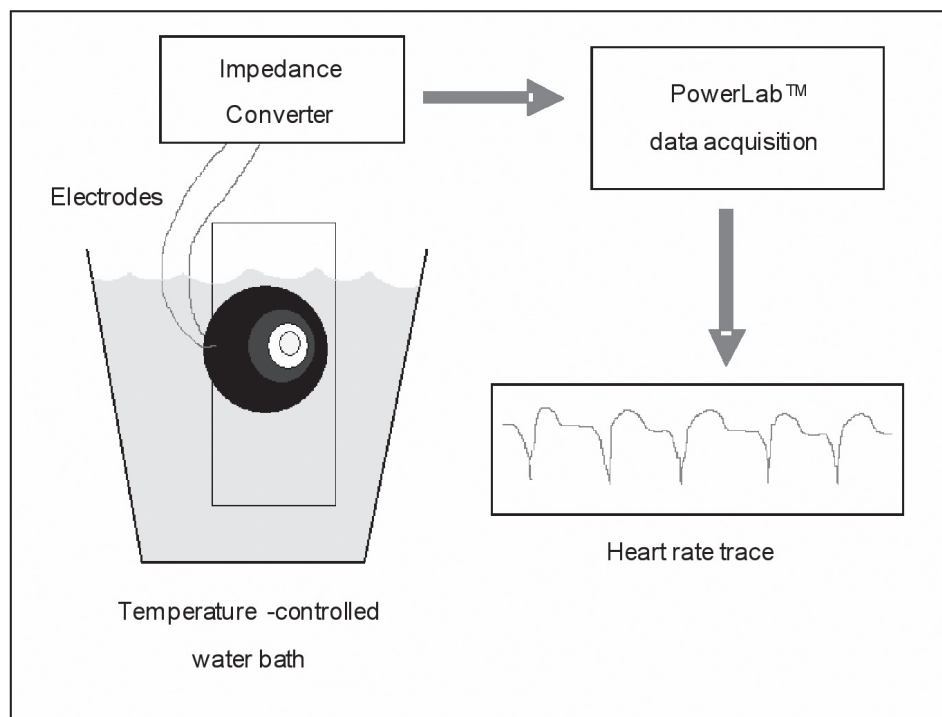Heart rates of snails from all three *Tegula* species increased in response to

Figure 2: Schematic diagram demonstrating design of equipment for temperature stress experiments.

heat stress, up to the ABT, after which they declined to zero (Fig. 3A). Once temperature was reduced, all species recovered cardiac function. Within 24 to 48 hours after experimentation, however, 92% of *T. funebralis* specimens and 100% of *T. brunnea* and *T. montereyi* specimens had died. ABTs were easily determined by generating Arrhenius plots for each specimen (Fig. 3B).

Significant interspecific differences in break temperatures were observed (p<.001) (Fig. 4). The average ABTs for field-acclimatized *T. funebralis*, *T. brunnea*, and *T. montereyi* were 31°C, 25°C, and 24°C, respectively (Table 1). The differences between *T. funebralis* and both subtidal congeners were statistically significant, but *T. brunnea* and *T. montereyi* did not have significantly different ABTs. These results establish a positive correlation between upper cardiac limits and maximum habitat temperatures in *Tegula* congeners. Flatline temperatures followed this trend as well. Average FLTs for *T. funebralis*, *T. brunnea*, and *T. montereyi* were 39°C, 32°C, and 33°C, respectively (Table 1). Flatline temperatures showed much less intraspecific variation than Arrhenius break temperatures, and all interspecific FLT differences were statistically significant (p<.0001).

Fig. 4. ABT and FLT values for three congeners of field-acclimatized *Tegula* subjected to heating. Numbers of specimens: *T. funebralis* (13), *T. brunnea* (11) and *T. montereyi* (10). Error bars are standard errors of the mean.

In response to cold stress, heart rates of all species showed gradual declines to zero. No clear break points could be identified, so only FLTs were determined and compared. *T. funebralis* maintained cardiac function down to 2°C, while *T. brunnea* and *T.*
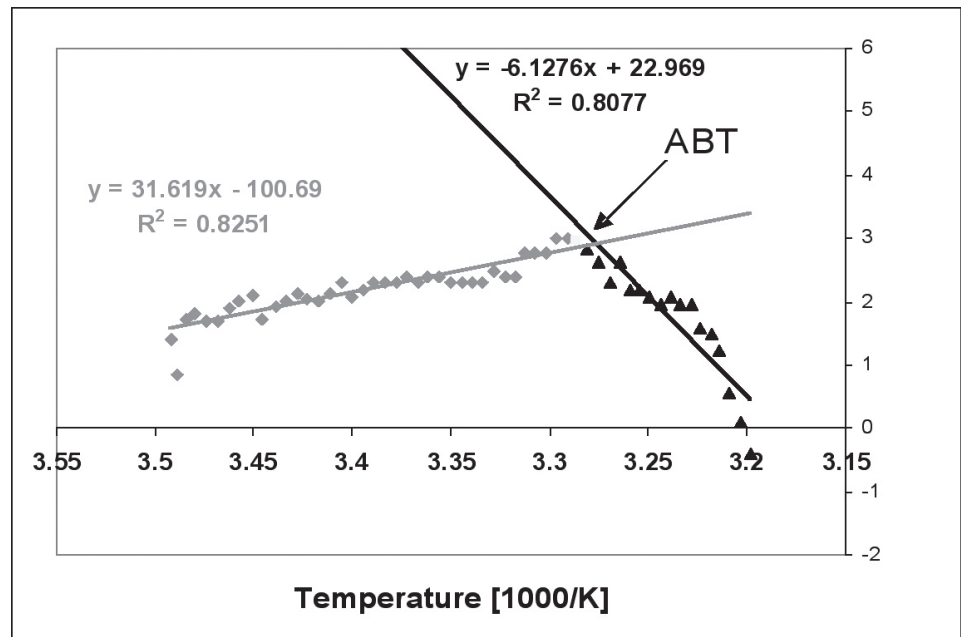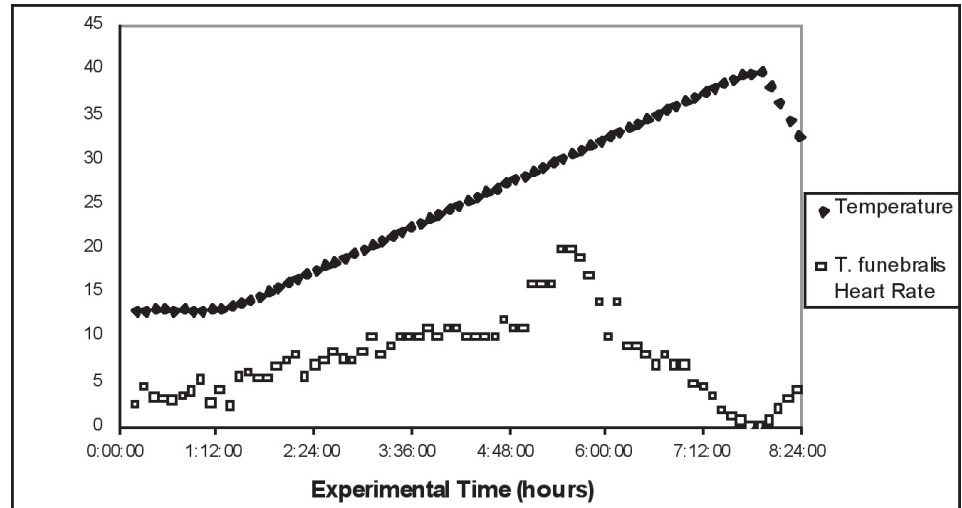

Figure 3: Effects of increasing temperature on heart rate of field-acclimatized T. funebralis. A) Typical response of heart rate to increasing temperature. B) Arrhenius plot of the same data. Inflection point was determined by visual inspection, and linear regression was performed on both sides of this point. ABT was calculated as the intersection of the two trendlines (see Materials and Methods).

montereyi hearts failed at 4°C and 5°C, respectively. *T. funebralis* can therefore endure significantly warmer and colder temperatures than its subtidal relatives. After all specimens had flat-lined, temperature was rapidly returned to ambient, and all species recovered heart function. Unlike the heat-stressed specimens, however, none of the cold-stressed snails died during the three-day holding period

following experimentation.

ABTs and FLTs were determined for 14°C and 22°C lab-acclimated snails of each species in response to heat stress. The two subtidal congeners, *T. brunnea* and *T. montereyi*, showed significantly higher ABTs after 22°C-acclimation than after 14°C-acclimation (Table 2). *T. funebralis* ABTs, however, did not significantly increase in response to warm-acclima-

tion. The 22°C-acclimated *T. funebralis* specimens had only slightly higher ABTs, on average, than the 14oC-acclimated snails (Fig. 5). For all species, flatline temperatures appeared to be less plastic than the Arrhenius break temperatures. The difference between 14°C and 22°C-acclimated FLTs was only approximately 1°C for all congeners.

Heart failure was not observed in the constant-temperature control run. Some heart rate fluctuation occurred in the first hour, but after this initial acclimation, cardiac function remained fairly stable in all the animals. During the 7 day holding period following the run, all control specimens were alive and functioning normally. These results indicate that the experimental set-up itself (drilling, electrode placement, and gluing) did not contribute to the heart failure and mortality observed in the temperature stress experiments.

Figure 4: ABT and FLT values for three congeners of field-acclimatized Tegula subjected to heating. Numbers of specimens: T. funebralis (13), T. brunnea (11) and T. montereyi (10). Error bars are standard errors of the mean.

Discussion

This study found a strong positive relationship between the upper cardiac thermal tolerance limits of *Tegula* congeners and their vertical position in the intertidal. In the two subtidal species, heart failure occurred at significantly lower temperatures than in *T. funebralis*, the higher-dwelling congener. A positive correlation was also found between vertical position and eurythermality. In other words, *T. funebralis* hearts proved capable of enduring the widest range of temperatures, while *T. brunnea* and *T. montereyi*, the lower-dwelling species, had narrower cardiac thermal tolerance ranges.

Comparing the average ABTs of field-acclimatized species to their maximum habitat temperatures exposes a critical finding. *T. funebralis* appears to be living much closer to its

Figure 5: Effects of thermal acclimation on Tegula congener Arrhenius break temperatures. Error bars are standard errors of the mean (n=6 for T. funebralis, n=5 for T. brunnea, and n=6 for T. montereyi).

upper cardiac limits than do its subtidal relatives. In their mid-intertidal habitats, *T. funebralis* individuals regularly reach temperatures above 30°C. Body temperatures as high as 33°C have even been reported [14]. Therefore, maximum habitat temperature of *T. funebralis* is actually higher than its

ABT (31.5º), meaning this species may experience compromised heart function on a regular basis. This is certainly not the case for the subtidal congeners. Maximum habit temperature for *T. brunnea* is approximately 23ºC, well below its ABT of 25.0ºC. Similarly, *T. montereyi* is unlikely to exceed 20ºC, which is far from its ABT of 24.2ºC.

Previous studies of *Tegula* thermal tolerance support my finding, that the higher occurring *Tegula* species is living closer to its thermal limits. For example, heat shock experiments have shown that mid- to high intertidal species such as *T. funebralis* and the subtropical species T. rugosa are likely to induce heat shock response more frequently than subtidal species such as *T. brunnea* and *T. montereyi* [14]. Activation of the heat-shock proteins occurs in response to thermal denaturation of proteins during heat stress [17]. The higher incidence of heat-shock response implicated in warm-adapted *Tegula* congeners may indicate more frequent protein denaturation.

Repairing such heat shock damage requires significant energy expenditures. Some research suggests that mid- to high-intertidal organisms may have higher 'costs of living' as a result [10]. Differential growth rates of *Tegula* congeners are consistent with this assertion. *T. montereyi* grows the fastest, *T. brunnea* is intermediate, and *T. funebralis*, the highest-occurring species, grows most slowly [18, 19]. Differences in heart rate observed in this study may also be related to energetic costs. *T. funebralis* heart rates were significantly higher than the two subtidal species, even when corrected for differences in body size. Perhaps an elevated cardiac output is necessary in *T. funebralis* in order to fuel the cellular processes associated with its frequent heat shock responses.

The results of the acclimation experiment may have great ecological significance, because thermal acclimation capacity can be used to predict a species' susceptibility to global climate change [20]. In this study, *T. funebralis* heart function showed the least acclimatory plasticity of all three congeners, since acclimation to 22ºC failed to induce a significant ABT increase in *T. funebralis*. Considering this result, and the finding that *T. funebralis* is living closer to its cardiac limits, it seems that *T. funebralis* may be at greater risk in the event of global warming than its subtidal congeners. Even a 1 or 2ºC increase in average global temperatures could physiologically compromise these animals.

Related studies on other marine invertebrate taxa raise similar concerns. Most notably, recent research on porcelain crabs of the genus Petrolisthes has shown that warm-adapted, eurythermal intertidal species may be more impacted by global warming than less heat tolerant congeneric subtidal species. Many physiological aspects of thermal tolerance, such as heart function and nerve function, have been studied comparatively in Petrolisthes congeners, and numerous significant correlations between vertical distribution and thermal limits have been found [9, 11, 20, 21]. As for cardiac function, the highest-occurring species, P. cinctipes and P. gracilis, have ABTs that are higher and closer to maximum habitat temperatures than the ABTs of low-intertidal and subtidal species. Also, cardiac limits and whole animal thermal limits of high-intertidal Petrolisthes species show lower

		T. funebralis (N=6)	*T. brunnea* (N=6)	*T. montereyi* (N=6)
Heat Stress	**ABT**	31.1± 0.7°C	25.0 ± 0.5°C	24.2 ± 0.7°C
	FLT	39.4 ± 0.2°C	32.4 ± 0.2°C	33.1 ± 0.1°C
Cold Stress	**FLT**	2.1 ± 0.2°C	3.5 ± 0.3ºC	4.8 ± 0.5°C

Table 1: Critical temperatures of field-acclimatized Tegula congenrs in response to heat stress and cold stress. Average Arrhenius break temperatures (ABTs) and flatline temperatures (FLTs) are listed, with associated standard errors of mean.

		T. funebralis (N=6)	*T. brunnea* (N=5)	*T. montereyi* (N=6)
ABT	14°C-acclimated	28.5 ± 0.5ºC	20.2 ± 0.8ºC	21.7 ± 0.8ºC
	22°C-acclimated	30.1 ± 0.5ºC	26.8 ± 0.7ºC	25.7 ± 0.6ºC
FLT	14°C-acclimated	39.8 ± 0.2ºC	31.7 ± 0.4ºC	33.6 ± 0.3ºC
	22°C-acclimated	40.7 ± 0.4ºC	32.8 ± 0.3ºC	34.1 ± 0.2ºC

Table 2: Critical temperatures of lab-acclimated Tegula congeners in response to heat stress. Average Arrhenius break temperatures (ABTs) and flatline temperatures (FLTs) are listed, with associated standard errors of mean.

acclimation capacities than lower-oc-curring congeners [20].

Thus, warm-adapted *Tegula* species as well as warm-adapted Petrolisthes species appear to be at highest risk in the event of global climate change. Although only a few genera are currently implicated, it does seem possible that this may be a broader trend in the marine intertidal zone: high-occuring intertidal organisms, in general, could be more susceptible to the effects of global warming.

One potential concern in this study centers on the fact that thermal limits were only determined for submerged specimens. All heart rate monitoring was performed in water, but *T. funebralis* is most likely to encounter heat stress when exposed to air. In all likelihood, however, the conclusions reached are still accurate, since temperature change in water is likely to be less stressful than temperature change in air. During emersion, snails may face additional stress from desiccation and restricted respiration, which might compound the effects induced by temperature alone. Therefore, the results of this study might actually underestimate the thermal tolerance limits of *Tegula* congeners, particularly *T. funebralis*. If this is the case, *T. funebralis* may be even more threatened by global warming than is suggested by these submersion experiments.

To expand upon this study, future experiments will aim to assess the differential effects of stresses in air versus water. The effects of thermal, dessication, and respiratory stresses could be tested in order to further elucidate the process of heart failure in intertidal organisms like *Tegula*. Additionally, other physiological parameters contributing to thermal sensitivity should be examined for *Tegula* congeners. Studying the effects of temperature on factors such as mitochondrial respiration, membrane fluidity, and nerve function could contribute to a more comprehensive understanding of heat stress and thermal tolerance.

Because of the potential effects of global warning on ecosystems around the world, pursuing an understanding of the physiological mechanisms involved in temperature stress is absolutely critical.

Acknowledgements

Literature Cited

1. Vernberg W. B. & Vernberg, F. J. (1972). Environmental physiology of marine organisms. New York: Springer-Verlag.
2. Denny, M. W. (1988). *Biology and the mechanics of the wave-swept environment.* Princeton: Princeton University Press.
3. Connell, J. H. (1961). The influence of interspecific competition and other factors on the distribution of the barnacle *Chthamalus stellatus*. Ecology, 42, 710-723.
4. Edney, E. B. (1961). The water and heat relationships of fiddler crab (Uca spp). *Trans. Roy. Soc. South Africa*, 34, 71-91.
5. Paine, R. T. (1969). The *Pisaster-Tegula* interaction: prey patches, predator food preference, and intertidal community structure. *Ecology*, 50, 950-961.
6. Hofmann, G. E. & Somero, G. N. (1995). Evidence for protein damage at environmental temperatures: seasonal changes in levels of ubiquitin conjugates and hsp70 in the intertidal mussel *Mytilus trossulus*. J. Exp. Biol., 198, 1509-1518
7. Wethey, D. S. (1983). Geographic limits and local zonation: The barnacles *Semibalanus* (Balanus) and *Chthamalus* in New England (USA). Biol. Bull., 165, 330-341.
8. Jensen, G. C. Armstrong, D. A. (1991). Intertidal zonation among congeners: factors regulating distribution of porcelain crabs *Petrolisthes* spp. (Anomura: Porcelanidae). *Mar. Ecol. Prog. Ser.*, 73: 47-60.

9. Stillman, J. H. & Somero, G. N. (2000). A comparative analysis of the upper thermal tolerance limits of eastern pacific porcelain crabs, genus *Petrolisthes*: Influences of latitude, vertical zonation, acclimation, and phylogeny. *Phys. and Biochem. Zool.*, 73, 200-208.

10. Somero, G. N. (2002). Thermal physiology and vertical zonation of intertidal animals: optima, limits, and costs of living. *Integ. and Comp. Biol.*, 42, 780-789.

11. Stillman, J. H. & Somero, G. N. (1996). Adaptation to temperature stress and aerial exposure in congeneric species of intertidal porcelain crabs (genus *Petrolisthes*): correlation of physiology, biochemistry and morphology with vertical distrubtion. *J. Exp. Biol.,* 199, 1845-1855.

12. Watanabe, J. M. (1984). The influence of recruitment, competition, and benthic predation on spatial distributions of three species of kelp forest gastropods (Trochidae: *Tegula*). *Ecology*, 65: 920-936.

13. Hellberg, M. E. (1998). Sympatric sea shells along the sea's shore: the geography of speciation in the marine gastropod *Tegula*. *Evolution*, 52, 1311-1324.

14. Tomanek, L. & Somero, G. N. (1999). Evolutionary and acclimation-induced variation in the heat-shock responses of congeneric marine snails (genus *Tegula*) from different thermal Habitats: implications fo limits of thermoterlance and biogeography. *J. Exp. Biol.*, 202, 2925-2936.

15. Tomanek, L. & Somero, G. N. (2000). Time course and magnitude of synthesis of heat-shock proteins in congeneric marine snails (genus *Tegula*) from different tidal heights. Phys. and Biochem. Zool., 73, 249-256.

16. Tomanek, L. & Sanford, E. (2003). Heat-shock protein 70 (hsp70) as a biochemical stress indicator: an experimental field tet in two congeneric intertidal gastropods (genus *Tegula*). *Biol. Bull.*, 205, 276-284.

17. Feder, M.E. & Hofmann, G.E. (1999). Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology. *Annual Review of Physiology,* 61, 243-282.

18. Frank, P.W. (1965). Shell growth in a natural population of the turban snail, *Tegula funebralis*. *Growth*, 29, 395-403.

19. Watanabe, J. M. (1982). *Aspects of community organization in a kelp forest habitat: factors influencing the bathymetric segregation of three species of herbivorous gastropods.* Ph.D. Diss., University of California, Berkely.

20. Stillman, J. H. (2003). Acclimation capacity underlies susceptibility to climate change. Science, 301, 65.

21. Stillman, J. H. (2002). Causes and consequences of thermal tolerance limits in rocky intertidal Porcelain crabs, genus *Petrolisthes. Integ. and Comp. Biol.*, 42, 790-796.

Emily Stenseng

Emily Stenseng is a senior majoring in Biological Sciences and a minor in chemistry. While at Stanford, she has spent quarters in Australia and at Hopkins Marine Station, pursuing her deep interest in marine biology. In addition to her scientific endeavors, Emily sings classically, and has performed in several operas and early music productions on campus. Next year she will return to her home state to begin veterinary school at Texas A&M University. She would like to thank her research advisor, George Somero, for all of his guidance and support. Additionally she extends her gratitude to Caren Braby, Maxine Chaney, and Jim Watanabe for their invaluable assistance with this project.

The Norms that Weren't

ASEAN's Shortcomings in Dealing with Transboundary Air Pollution

Bryan Tan

In this paper, I examine the shortcomings of the Association of Southeast Asian Nations (ASEAN) as a treaty organization. ASEAN has created an international community for dealing with environmental disasters, most notably the Southeast Asian haze. I use the contextual, economically focused background of ASEAN as a means to understanding its present-day structure, which I conclude is inefficient in dealing with non-economic issues. I then discuss some possible changes that might be made to address ignored threats of the Southeast Asian haze.

The Southeast Asian smoke haze of 1997, which resulted from widespread and synchronized incidents of incendiary land-clearing agricultural practices in Indonesia, is described by many experts as one of the century's worst environmental disasters. Leaving a charred area the size of Costa Rica utterly devastated with reported damages amounting to US$4.5 billion,[1] the haze cast a poisonous shadow over 70 million people living in Malaysia, Singapore, Brunei Darussalam, The Philippines, Thailand, and Indonesia itself.[2] Globally localized but doubtlessly a transnational environmental issue, the corresponding economic downturn and ecological damage, coupled with the astronomical implications for human health, was unprecedented – at its peak, the haze even forced the closure of several Southeast Asian international airports due to the incredibly poor visibility. At the time a Singaporean high-schooler, I remember looking out of my window one day and barely being able to

make out the outline of the family car, which was parked on the street a mere twenty meters away from my bedroom. As a further demonstration of the extent of the catastrophe, biodiversity was also harmed: endangered species including rhinos, tigers and orangutans were driven precariously closer to extinction.[3]

Interestingly, despite these more immediately obvious ramifications, what many scholars and political scientists found most unsettling about the event was the fact that this haze was not an isolated phenomenon. In fact, Southeast Asia had been plagued by periodical occurrences of haze since 1982 (with the most recent haze outbreak occurring in 2004). Although the haze of 1997 was greater in severity of effect, reach and consequence, it was no different in terms of origin and cause than the hazes of 1994, 1991, or even 1986;[4] all this in spite of several agreements signed after the first haze occurrence by the Association of Southeast Asian Nations

(ASEAN) committing member nations to cooperate in preventing transboundary air pollution. Clearly, ASEAN was not doing enough to enforce its environmental policies. In this paper, I will argue that the main reason behind this apparent ignorance is an internal conflict of priorities: ASEAN's primary focus as a regional treaty organization is not and has never been tackling environmental issues. Because of this, the present-day organizational structure of ASEAN ignores many general guidelines of international environmental lawmaking in the name of economic advancement and is ill-suited to handle large-scale environmental disasters like the recurring Southeast Asian haze.

To understand the haze's suffocating extent, and to simultaneously analyze the shortcomings of ASEAN in handling the matter, it is first necessary to comprehend the scientific nature and complex causal origins of the haze itself, as well as to consider the

organizational structure of ASEAN.

Definition of Haze

The ASEAN Haze Action website, which was officially spawned as a result of ASEAN's Regional Haze Action Plan of 1997 and contains among other utilities an intranet linking ASEAN meteorological services and environmental agencies to improve communications and enhance the effectiveness of existing early warning and monitoring systems,[5] defines transboundary haze as:

> Sufficient smoke, dust, moisture, and vapor suspended in air to impair visibility. Haze pollution can be said to be "transboundary" if its density and extent is so great at source that it remains at measurable levels after crossing into another country's air space.[6]

The damage done to both the economy and people of Southeast Asia is also elucidated on the website in terms of the detrimental effects particulate matter exerts on those exposed:

> Particulate matter [that comprise haze] less than 10 micrometers in size … can penetrate deep into the lungs. In recent studies, exposure to particulate pollution— either alone or with other air pollutants—has been linked with premature death, difficult breathing, aggravated asthma, increased hospital admissions and emergency room visits, and increased respiratory symptoms in children.[7]

It is easy to see why haze poses such a huge threat to the health of those living in its sphere of influence. As such, the haze affected not only Southeast Asian natives, but also the tourism industry, which has traditionally been one of the core tenets of the regional economy. On top of this, the Regional Haze Action Plan lists a variety of other sectors that were adversely affected, namely air, water and land transport, shipping, construction, forestry and agriculture.[8]

Causes of the Haze

With the multitude of negative implications haze carries, the crux of the issue for ASEAN must have surely concerned learning from its mistakes and working to prevent the man-made forest and grass fires that have, in a regular cycle, given rise to this problem. It is difficult to believe that hydrated rainforest vegetation can burn at all in a tropical area like the Indonesian state of Kalimantan, the locality held responsible for the causal fires. Geographically centered on the equator, the forests of Kalimantan experience remarkably high levels of rainfall, resulting in a continuous saturated state of atmospheric humidity which is seemingly impervious to burning of any kind. However, a combination of factors has contributed to the susceptibility of these forests to burning, most notably the El Niño Southern Oscillation (ENSO), which in every four years or so culminates in a Southeast Asian dry season,[9] and various human activities, such as rapid deforestation and the draining of peat swamps for rice cultivation.[10]

Government revenue in Indonesia relies heavily upon its timber industry. The granting of excessive timber concessions to logging companies and the poor enforcement of selective deforestation policies, largely due to corrupt Indonesian government officials, leads to extensive forest cover damage and a forest floor littered with brittle timber debris, which exacerbates the drying of forests in seasons with low rainfall.[11] The fires themselves are also started through human activity. When evaluating human responsibility, it is important to draw the distinction between fires caused by native forest communities and those incited by large commercial operations.

The Dayak tribe, an indigenous Kalimantan forest people, has been— for thousands of years before the first ever incident of transboundary haze in Southeast Asia—practicing shifting cultivation, also known as slash-and-burn agriculture. This entails the controlled use of fire to clear land for their agricultural subsistence, and upon exhausting the vegetative nutrition of a particular plot of land, moving to a different plot and repeating the process.[12] The Dayak culture emphasizes the necessity of both a deep respect for nature and harmony with the natural environment. They can be exempted from blame not only because the small-scale pollution they produce is relatively negligible and not an international environmental issue but also for the time it takes for a plot of land to fully lose its agricultural value: typically two to four years. The Dayaks' centuries of experience shows in that during the cropping of a new plot of land, the previous plot is left to fallow and regenerate its sustenance, sometimes for up to 60 years. The short cropping period, followed by a long fallow, restores carbon and nutrients to the system and facilitates the reinvasion of negatively impacted flora and fauna.[13]

This process lies in stark contrast with the techniques applied by the burgeoning timber and oil palm

industries. To maximize profits and with minimal regard for long-term damage, large sections of rainforest are burned down and converted to plantations. Burning is the cheapest method of land-clearing known to relatively impoverished Indonesia, and since the aforementioned industrial programs receive extensive support from the Indonesian government, companies are in effect given the go-ahead to exploit Indonesia's natural resources to their financial advantage.

Goals and Structure of ASEAN

Understanding the nature of haze helps to demonstrate why it is a problem to which ASEAN should assign a higher priority. Understanding the causes of haze in relation to the political structure of Indonesia and the organizational structure of ASEAN should help to demonstrate why ASEAN has been so ineffective in curbing transboundary air pollution in Southeast Asia. The Treaty of Amity and Cooperation, signed at the First ASEAN Summit in 1976, glorifies within its fundamental principles the following ideals:

a) Mutual respect for the independence, sovereignty … and territorial integrity of all nations;

b) The right of every state to lead its national existence free from external interference, subversion or coercion;

c) Non-interference in the internal affairs of one another.[14]

These ideals are well suited and indeed integral to ASEAN's primary purpose, which its founding document, the Bangkok Declaration of 1967 (also known as the ASEAN Declaration) clearly announces as:

To accelerate the economic growth, social progress

and cultural development in the region through joint endeavors in the spirit of equality and partnership in order to strengthen the foundation for a prosperous and peaceful community of Southeast Asian nations…[15]

When ASEAN was founded as an international conglomeration in 1967, its main focus was to strengthen economic relations among its members. With ten current member nations (Brunei Darussalam, Cambodia, Indonesia, Laos, Malaysia, Myanmar, The Philippines, Singapore, Thailand and Vietnam), ASEAN has a combined population of around 500 million spread over about 4.5 million square kilometers of land. The member nations of ASEAN have today come to define the geographical limits of "Southeast Asia," and fuelling them is a combined GDP of US$737 billion and an overall trade of US$720 billion which has transformed the region into what Kofi Annan, Secretary-General of the UN, terms an "indispensable and real force to be reckoned with even far beyond the region."[16]

With development and a solidifying of organizational identity comes greater shared responsibility, and today's ASEAN, as a treaty, an international organization, and a geographical region, is responsible for a variety of other concerns, including protecting the Southeast Asian environment. However, a detailed study of ASEAN's founding documents, including the Bangkok Declaration and the Treaty of Amity and Cooperation, reveals no mention whatsoever of environmental issues or policy; in fact, a cursory Internet Explorer word search turns up not a single appearance of the word "environment".[17] This

is surprising because almost every other conceivable social, economic and developmental issue is addressed in these two treaties, including some which do not necessarily supersede the environment in terms of absolute global importance, such as the promotion of Southeast Asian studies.[18] Of course, this does not imply that separate environmental accords have not been signed since the founding of ASEAN 38 years ago, but it does mean that the original intentions of ASEAN were biased towards very different issues. History shows that the environment came into the picture much later, and only as a result of ASEAN realizing that, as an organization growing in international stature and repute, adopting regional environmental policies would become inevitable.

Shortcomings of ASEAN in Dealing with the Haze

It is no secret that environmental concern and industrialized development often stand in direct opposition. In a region like Southeast Asia, with developing countries making up eight of the ten member nations (only Brunei Darussalam and Singapore are not classified by the World Bank as "developing"),[19] economic advancement has always, albeit understandably, taken precedence over environmental concerns. Only when the haze issue was recognized as a serious threat to the regional economy did ASEAN respond with its Regional Haze Action Plan. Since then, the issue has appeared in two other agreements: the Jakarta Declaration on Environment and Development of 1997[20] and the ASEAN Agreement on Transboundary Haze Pollution of 2002.[21] Tellingly, all of these treaties

fail to take into consideration one signficant factor: the conflicting political culture of individual nations – remember, this is a treaty organization that received widespread global criticism for not taking incisive action against the genocidal regime of Cambodia's Pol Pot.[22] Perhaps this is a function of ASEAN's outdated fundamental principles,[23] but in any case, it is imperative to recognize that environmental agreements must be based on certain norms, which are undefined in the case of Southeast Asia.

Corruption is rife in Southeast Asia, and among the most serious offenders is Indonesia. NGO Indonesia Corruption Watch spokesperson Luky Djani explains, "Corruption in Indonesia is ingrained [in the culture] and systematic."[24] In Indonesia, bribery and use of government funds for personal benefit are ways of life. With such a mindset, it is extremely difficult to work around the ideals of sustainable development, intergenerational equity and common heritage of humankind,[25] simply because they are not thought of as norms. What is assumed to hold true for most rational, law-abiding states amounts to nothing within the radically differing political realities of ASEAN. In this aspect, it is almost inevitable that the current haze situation has not progressed beyond the way it was in 1982. One of the Regional Haze Action Plan's primary objectives is to prevent land and forest fires through better management policies and enforcement,[26] but how is this to be observed through multilateral pressure when the defining value of ASEAN is non-intervention?[27]

Having studied the background of the recurrent haze crises, it becomes apparent that their root cause is firmly embedded in the Indonesian system,

and merely agreeing to regionally enforce the agreements without any actual action does not solve anything because international law cannot be effective without first establishing a general acceptance of customary norms.[28] Precisely because of this, ASEAN nations struggle to draw the line between respecting their neighbors' privacy and cooperatively mediating a situation which ultimately affects the entire region. Even the most comprehensive of the haze treaties, the ASEAN Agreement on Transboundary Haze Pollution, which claims to provide a legal framework in addressing the problem, lists a number of internal solutions which rely on a country's own resources to police.[29] To this day, ASEAN has yet to implement any form of serious punishment against Indonesia – a stance attesting to the primary goals of the organization. To ASEAN, actions such as imposing trade sanctions would be unthinkable; encouraging free trade and increasing economic prosperity are still paramount. For now, the environment takes a backseat.

Possible Rectifications

Solving the Southeast Asian haze issue is a daunting prospect. The challenge of sustainable development in Indonesia has not been handled well, considering the reckless disregard for the long-term future of the country's natural resources. To change this, corruption in Indonesia must first be addressed, preferably by an international organization such as the UN, which may have enough global backing and clout to make its point firmly. Currently, even states and NGOs that want to offer aid to the tsunami victims in Indonesia are hesitant, because they cannot be sure

where their money will end up. Newly elected Indonesian President Susilo Bambang Yudhoyono, who took office in September 2004, is trying to rectify this, and has promised a campaign against corruption, which has sadly been met with an overwhelming feeling of skepticism.[30]

ASEAN must redesign its treaty structure regarding environmental issues so as to better the serve the dynamic, ever-evolving interests of the region. Its principle of non-intervention needs to be reinvented to allow for transboundary environmental policy enforcement, simply because Indonesia does not have the finances or technology to monitor itself in this department, nor does it care to do so. If the UN can provide a certain amount of professional anti-corruption advice and if ASEAN can undergo a profound shift in its mindset, then the source of the problem will be addressed and the occurrences of transboundary air pollution in Southeast Asia will be, if not totally eradicated, at least less common.

Conclusion

The sad truth is that there exists in certain parts of our world communities for which starting large-scale forest fires in the name of land clearing and industrial advancement is an accepted, or at least tolerated, agricultural practice. If this practice is seen as the norm in Indonesia, then it is pointless to sign treaty after treaty calling for the country to clean up its act. Treaties are based on a common normative understanding, and in the case of the Southeast Asian haze, active multilateral enforcement of environmental policy is required.

ASEAN as a treaty organization has been successful

in certain aspects—the economy of Southeast Asia has grown spectacularly since its inception, and is progressing well towards the final destination of development—but it is miserably ineffective in addressing the problem of transboundary air pollution. Once the ASEAN leaders begin to realize that they are at a stage of their development at which more concern can and should be paid to issues other than economic development, they will hopefully recognize how they have been failing the future generations of Southeast Asians and change their policies. ASEAN indeed covers a limited geographic scope, but transboundary air pollution is not contained by treaty affiliation, and once ASEAN awakens to its harsh reality, the global community will be better off for it.[1]

Endnotes

[1] Indonesia's Fire and Haze: The Cost of Catastrophe (David Glover et al. ed., 1999). Excerpt available at <http://web.idrc.ca/en/ev-9410-201-1-DO_TOPIC.html> (visited Jan. 17, 2005).

[2] "Effects of Indonesian Forest Fires," James R. Lee, American University Trade Environment Database, available at <http://www.american.edu/projects/mandala/TED/indofire.htm> (visited Jan. 15, 2005).

[3] Indonesia's Fire and Haze, supra note 1.

[4] Effects of Indonesian Forest Fires, supra note 2.

[5] ASEAN Regional Haze Action Plan, Dec. 23, 1997, available at <http://www.aseansec.org/9059.htm> (visited Jan. 15, 2005).

[6] Haze Info, ASEAN Haze Action Online, available at <http://www.haze-online.or.id/help/firehaze.php?PHPSESSID=52a7d97c157e3336622e16cbd3037a13> (visited Jan. 15, 2005).

[7] Id.

[8] ASEAN Regional Haze Action Plan, supra note 5.

[9] "The Haze over Southeast Asia: Challenging the ASEAN Mode of Regional Engagement," James Cotton, Pacific Affairs, Vol. 72, No. 3, Aut. 1999, available at <http://links.jstor.org/sici?sici=0030-851X%28199923%2972%3A3%3C331%3AT%22OSAC%3E2.0.CO%3B2-0> (visited Jan. 16, 2005).

[10] "Politics and Peat: The One Million Hectare Sawah Project," Inside Indonesia, no. 48, Oct.-Dec., 1996, available at <http://www.insideindonesia.org> (visited Jan. 17, 2005).

[11] The Haze over Southeast Asia, supra note 9.

[12] Effects of Indonesian Forest Fires, supra note 2.

[13] "Alternatives to Slash and Burn Agroforestry, Shifting Cultivations or Swidden Farming," Erick C.M. Fernandes, 1998-1999, Cornell University Crops and Soil Sciences International Agroforestry Resources, available at <http://www.css.cornell.edu/ecf3/Web/new/AF/ASB_01.html> (visited Jan. 17, 2005).

[14] Treaty of Amity and Cooperation in Southeast Asia, Feb. 24, 1976, available at <http://www.aseansec.org/1217.htm> (visited Jan. 16, 2005).

[15] Bangkok Declaration, Aug. 8, 1967, available at <http://www.aseansec.org/1212.htm> (visited Jan. 17, 2005).

[16] Overview: The Association of Southeast Asian Nations, available at <http://www.aseansec.org/64.htm> (visited Jan. 14, 2005).

[17] Bangkok Declaration and Treaty of Amity and Cooperation in Southeast Asia, supra notes 14 and 13.

[18] Id.

[19] World Bank Group Data & Statistics, available at <http://www.worldbank.org/data/countryclass/classgroups.htm> (visited Jan. 18, 2005).

[20] Jakarta Declaration on Environment and Development, Sep. 18, 1997, available at <http://www.aseansec.org/6085.htm> (visited Jan. 18, 2005).

[21] ASEAN Agreement on Transboundary Haze Pollution, Jun. 11, 2002, available at <http://www.aseansec.org/6086.htm> (visited Jan. 18, 2005).

[22] The Cambodia Daily ASEAN Supplement, available at <http://www.camnet.com.kh/cambodia.daily/asean/11.htm> (visited Jan. 18, 2005).

[23] Treaty of Amity and Cooperation in Southeast Asia, supra note 14.

[24] "Corruption in Indonesia is Worrying Aid Groups," Raymond Bonner, The New York Times, Jan. 13, 2005, available at <http://www.nytimes.com/2005/01/13/international/worldspecial4/13corruption.html> (visited Jan. 18, 2005).

[25] The Origin and Emergence of International Environmental Norms, Armin Rosencranz, Hastings International and Comparative Law Review, Vol. 26, No. 3, Spring 2003.

[26] ASEAN Regional Haze Action Plan, supra note 5.

[27] ASEAN and the Principle of Non-Intervention: Practice and Prospects, John Funston, Institute for Southeast Asian Studies Journal, No. 5, Mar. 2000.

[28] International Environmental Law and Policy (2nd Ed.), David Hunter et al., 2002.

[29] ASEAN Agreement on Transboundary Haze Pollution, supra note 21.

[30] Corruption in Indonesia, supra note 24.

Bryan Tan

Mentor: Professor Armin Rosencranz

Bryan Tan is a sophomore from sunny Singapore majoring in International Relations (Comparative International Political Economy) and Communication (Public Affairs) with a Drama minor (Acting). An award-winning singer, actor and public speaker, he sings with the Stanford Harmonics and has to date appeared in over ten Stanford productions for the Ram's Head Theatrical Society, Drama Department and Stanford Film Society. With a strong interest in Southeast Asian issues, Bryan has also had an article about Singaporean media censorship published in CommunicAsians magazine. Readers should know that Winter 2005 marks the first time in four years that he has had a head of fully black hair.

Married Women and AIDS Vulnerability
Voices from Rural South Africa

Jenny Tolan

This essay examines the specific vulnerability of married women to AIDS, using voices of women from rural South Africa. Studies are beginning to reveal that married women are at a higher risk for AIDS than single women. This essay finds that three major factors contribute to this vulnerability: migrant labor, lobola – or bride price - and gendered economic inequality. These factors lead to heightened AIDS vulnerability by lowering the agency of married women and by increasing their sexual risk behavior. This essay argues that current AIDS prevention policies prove largely ineffective when applied to married women. To effectively combat AIDS, female-controlled prevention methods should be placed at the forefront of research efforts.

South Africa, Africa's wealthiest nation, faces one of the fastest growing rates of HIV/AIDS in the world. In the past 15 years, the nation has seen a huge jump in HIV prevalence, from one percent of the population in 1990 to 20 percent in 2001.[1] In 2004 these figures reached over 26 percent, making South Africa the nation with the world's highest rate of HIV/AIDS infection.[2] But HIV/AIDS does not affect all populations equally. As in most of sub-Saharan Africa, women in South Africa are 30 percent more likely to be HIV positive than men.[3] Even more surprising, AIDS does not target all women consistently. Research has found that married women and women in long-term monogamous relationships run a greater risk of contracting HIV than non-married women.[4] Because most cultures and policies encourage monogamy and commitment in the face of AIDS, these statistics are shocking. Why are women in South Africa, and particularly married women, more susceptible to HIV/AIDS than men? What conditions of South African marriage make women more vulnerable to the effects of this global pandemic?

I argue that rural, married women in South Africa have specific vulnerabilities to HIV/AIDS as a result of three major factors: migrant labor, lobola, and gendered economic inequality. These three factors lead to heightened AIDS vulnerability in two ways: they lower the agency of women and they increase sexual risk behavior. Women experience lowered agency in that they have virtually no ability to refuse sex or to demand the use of condoms. They experience increased sexual risk behavior in that they have higher coital frequency, decreased condom use, and exposure to partners with higher rates of infection. Although women are also twice as biologically vulnerable to HIV as men, it is their lack of power and high risk behavior that has the largest impact on their HIV risk.[5] The danger of this vulnerability lies not only in its impact on women, but also on the population at large as women pass the disease onto their children.[6] To conclude I argue that current AIDS prevention policies prove largely ineffective when applied to married women.[7]

Up to this point, little research has focused specifically on married women and HIV/AIDS. One study that looked specifically at marriage and HIV risk was published in September of 2004 by Shelley Clark. This study looked at adolescent girls and came up with the shocking results that—among girls aged 15 to 19—being married was associated with an increase of greater than 75 percent in the odds of being HIV-positive compared with the odds for sexually active unmarried girls.[8] This statistic underscores the urgent need for further research on married women and AIDS.

In the summer of 2004, I spent eight weeks at the Masoyi Home Based Care Project in Mpumalanga, South Africa. The Masoyi tribal area is a rural district home to 220,000

black Africans, most of whom live in extreme poverty. The unemployment rate is nearly 75 percent, and approximately 32 percent of the adult population—an estimated 132,000 people—are HIV positive.[9]

During my stay I interviewed ten women and two men who work with the Masoyi program. All of my subjects were either married or in long-term relationships. Though the interviews did not reveal the HIV status of my subjects, the common themes that arose from their stories expose this group's particular vulnerabilities to HIV/AIDS and lead to a broader discussion of gender, marriage and disease in South Africa.

The Factors of Vulnerability
Migrant Labor

The system of migrant labor leads to high sexual risk behavior for women by increasing their exposure to partners with high rates of infection.[10] The migrant labor system is a consequence of overwhelming unemployment in rural areas where men seeking jobs are forced to travel to the mines or cities. The women of Masoyi described this scenario in their own words:

Jabulile:
My husband is working in Witbank, welding some broken things. He is working in a workshop. He has been there a long time – he got this job in 1987 – when I got the first child. He comes home month end. He says he's not enjoying working there, but because there is no work here, he is working there.

While men are away from home, it is common for them to have multiple sexual partners.[11] This is a result of the conditions of their work

Jabulile (Joy) – Age 35 - Orphan Coordinator of Masoyi Home Based Care Project, at right with with Yandisua – a child of one of the orphans

as well as the cultural acceptance of male infidelity. First, workers' housing in the mining and urban industries is often unbearable. Men live in single-sex barracks, often twelve to sixteen per room, with little space and no privacy.[12] Studies show that men say they can not stay celibate while separated from their wives for such long periods of time.[13] Jabulile describes the norm of separation from her husband:

Jabulile:
Since I was married to him – we didn't stay together for a long time. He stay for three days and go – stay for three days and go. If there was work here for him to work every day and come home, he would stay. But there is none – And he's not somebody who is educated so he can find any job.

The practice of male polygamy is also culturally accepted by most men and women in South Africa. In a 2003 study by Ribiero Da Cruz about HIV and condom use in South Africa, the male participants described having multiple partners as acceptable and even desirable.[14] Of the subjects interviewed, 59 percent of males had more than one sexual partner, and 22 percent had three or

more sexual partners.[15] Most of my subjects were aware of this risk:

Jabulile:
About sickness – I'm worried about that. Because I know my husband – he is not someone who is faithful. Ah, I'm really worried.

Migrant men can be involved with women at work in several ways: engagement with prostitutes, short-term casual relationships, or long-term relationships with 'second wives.'[16] The highest risk comes with prostitution. Commercial sex industries are rampant in mining and industrial areas. Caldwell et al. write, "In most of Africa it is rural migrants to the city who are most likely to be found in the slum and shanty town bars where the commercial sex workers are more likely to have uncured STDs and are probably more likely to have many different customers and be HIV positive."[17] The culture of poverty and prostitution leads men to high exposure to STDs and HIV.

In 1993 it was already estimated that nearly half of the mine workers returning to rural areas after work were infecting their wives and other women.[18] Since then, AIDS rates in South Africa have skyrocketed,

and so have the levels of infection of migrant workers. Caldwell et al. write, "With the exception of [mother to child and infected blood transmission] the HIV levels in rural areas may be almost entirely the result of persistent reinfection brought back from the towns by returning migrants."[19] Because migrant workers generate income, they are more likely to be able to afford brides and are therefore more likely to be married.[20]

Lobola

The custom of paying a bride's family in exchange for marriage is a lingering element of traditional patriarchy. Lobola as a tradition dates back to the founders of the Zulu nation. In the past century, however, it has shifted from traditional symbolism to a commercial practice often paid in cash.[21] As the marriage transaction has grown more commercial, the woman has become more like the property of the man.[22] Today, the payment of lobola is extremely widespread. Likhapha Mbatha of the University of Witwatersrand studied women in three South African provinces and found that bride price had been paid for 98% of the wives.[23]

The tradition of lobola lowers female agency by creating a sense of male ownership that leaves the wife subject to her husband's demands. Von Kapff writes, "No wife will dare to oppose her husband as she would be sent home, and her father would have to return most of the cattle."[24] Though divorce is accepted in South Africa today, a woman is responsible for repayment of the lobola to her ex-husband. This is largely impossible in poor areas when the woman's family has already 'eaten' the lobola.[25]

Lobola payment gives the husband full rights over his wife's productive and reproductive capabilities.[26] According to Barbara Klugman, the director of the Women's Health Project at the University of Witwatersrand, "If a husband initiates sex, his wife may not refuse him."[27] This standard also applies to condom use. According to customs of traditional marriage, a woman may not use contraceptives without the consent of her husband.[28] Unfortunately, studies have shown that women have a greater desire for condom use than their male counterparts.[29] For men in South Africa, a deep stigma surrounds the use of condoms. One subject in the study by Da Cruz stated "My boyfriend says using a condom is like eating a sweet with the wrapper on it."[30] Because husbands do not like to use condoms, the women are forced to agree.

Jabulile:

At one time I didn't want my husband to come to me without a condom. In that moment he saw that I was cross, so he accept the condom – he accept it. But as time goes on he say "I'm tired. I'm tired – I've paid lobola for you – you're my wife." And – there was a fight. In the end – he wins the fight. We don't use the condom. At the moment – I'm just afraid – I don't know my status.

As Jabulile's story makes clear, lobola can reaffirm male sexual dominance and deny women the right to request condoms. A shocking study from Zambia found that only 11 percent of women believed they had the right to ask their husbands to use a condom—even if he had proven himself to be unfaithful and was HIV-positive.[31]

In addition to lowering a woman's agency, lobola also contributes to HIV risk by increasing the sexual risk behavior of women. The first example of this is simply higher coital frequency. A study in Kenya and Zambia found that married girls have unprotected sex much more often and have been engaged in sexual activity for a longer period of their lives than have unmarried girls.[32] Considering that many partners are HIV positive, the mere reality of coital frequency is a sexual risk behavior.

A second instance of sexual risk behavior involves condom use. Lawson writes, "In rural communities social control, particularly over the sexual behavior of women, is strong because women are regarded as childbearers, whose duty it is to perpetuate the lineage of the husband's family."[33] With this high emphasis placed on producing children, Schoepf writes, "Couples that have not reached their desired family size will reject condoms, even when one spouse is HIV positive."[34] This reality compounds the HIV risk for married women.

A final factor of sexual risk behavior stemming from lobola is the tendency for women to be 'bought' and married by men who are significantly older and therefore able to pay. At least two studies, by Gregson et al. and Kelly et al., have shown that the age of a woman's partner is a major risk factor, and that having an older partner substantially raises HIV rates among adolescent girls.[35]

Despite these connections of lobola with gender inequality and AIDS, a contrasting viewpoint argues that lobola can actually benefit women. A thirteen year old girl told me in an interview that she did not like men having ownership over their wives. When I asked her why the community did not end the practice of lobola, however, her response was strong:

Left to Right: Thanduxolo – Age 13, daughter of Masoyi Project Coordinator; Busisiwe (Blessing) – Age 38 - Preschool Teacher, Masoyi Home Based Care Volunteer; Mumsy – Age 36 - Childcare worker for Masoyi Home Based Care Project, with her children outside her two room house

Thanduxolo:

We cannot get rid of lobola! If there was no lobola, men could just go from wife to wife. Once the man pays lobola he has given his money so he will stay with that wife. He has paid for her, so he must stay. Without lobola, men would never stay with one wife. They could just go.

In this way, the tradition of lobola may actually give women some social power. Scholar Von Kapff writes, "The more cattle paid, the better the marriage seems to work in the long-term, as the bridegroom frequently has to save over several years for the lobola and therefore chooses his bride carefully."[36] For these reasons, many South African women want to keep the custom alive.

This dichotomy surrounding lobola creates an extremely difficult situation. A policy solution would have to confront the gendered norms while somehow preserving the positive components of lobola that are part of African cultural tradition.

Economic Inequality

The final way that married women are at high risk is through their economic dependence on men. Rural African women form the majority of the poorest of the poor in South Africa with an average income of between R400 and R700 ($64-$112) per month.[37] In an area like Masoyi where the majority of those employed are migrant workers, women are unlikely to find employment. The mining industry, for example, employs 97 percent men and only three percent women.[38] Furthermore, while men are away for long periods of time, it is the woman's duty to stay at home to care for the family. Busisiwe speaks of her lack of money and subsequent reliance on her husband:

Busisiwe:

I need money. I want enough money to send my children to school. I get R280 ($45) for volunteering. My husband he is working at Kinrose mine. Do you know this mine? It is a gold mine. I think he is getting R1000 ($160) a month.

This economic reality leaves women little agency to stand up to men in fear of losing financial support. Catherine Albertyn writes, "The prevalence of women seeking sexual relationships to ensure food and shelter for themselves and their families is a widespread consequence of gendered poverty and inequality in South Africa."[39] This situation was common among women of Masoyi.

Mumsy:

In my family I was suffering. My husband – he was at Joburg since March – not coming back. No money to give me to buy some food. So my children are suffering. My husband is working at Kinrose mine. Since 1988 until today – it's a long time. He is making good money there, but he gives me little money. He is getting R1200 or R1300 (~$210) and he is giving me R400 (~$66) – is a little money – to buy some food.

Mumsy's words present a classic case of a migrant husband who refuses to give his wife a fair share of income. Already she lives in the most basic conditions, but at least if she stays with her husband, there is the possibility he will one day use his money to build a third room on her house. If Mumsy's husband does not

want to use condoms, she has little agency to demand otherwise.

Mumsy:
Ohhh! My husband doesn't like the condom. He doesn't like! So I'm scared about this. I advise him long time, but he say condom is a plastic. He doesn't like, he doesn't use. I say use a condom, he say no. I don't know what I can do.

Kalipeni et al. affirm, "Women are not always in the economic position to say no to partners who will not assent to using condoms."[40] This is true regardless of women's knowledge of HIV risk. The women of Masoyi were all aware of the risk of AIDS, but they were more concerned with the immediate necessity of feeding their children. The most common question surrounding married women and AIDS was the following: "If it's a choice between leaving a man to lower the risk of AIDS and putting food on the table for your children, what would you choose?"[41] The answer for most women was simple: stay with your husband, accept his demands, and pray you do not fall victim to AIDS.

Implications for Policy and Female-Controlled Prevention

In view of all of these conditions, the current policies for HIV/AIDS prevention prove widely ineffective for married women. Today's dominant global strategy is called the "ABC Strategy," and stands for "Abstain, Be Faithful, Use Condoms." Apparently successful in Uganda, this policy was adopted by the United States for the Bush Administration's global AIDS effort.[42] Unfortunately, married women cannot abstain from sex, cannot control their husbands' faithfulness, and cannot demand the use of condoms. Today's relatively high AIDS rates for married women may in fact be a sign that prevention methods have been effective for single women. To target married women, however, an innovative solution must bypass the control of the husband and give the woman agency.

The most promising strategy to decrease women's HIV risk is 'female-controlled prevention.'[43] One option that exists today is the female condom. Though a condom would still prevent procreation and might be problematic, research by UNAIDS in Costa Rica, Indonesia, Mexico and Senegal found that men are more willing to accept female condoms than male condoms.[44] Female participants in a study by Susser and Stein felt they could not insist that men used condoms, but seemed more confident that they could use a female condom.[45]

A second method with enormous potential is the microbicide. A microbicide is a spermicide-type gel that women could apply before intercourse to protect against HIV and STDs.[46] This option would give women full agency over their HIV risk because they could use the gel without consulting their partners.[47] There are currently 40 microbicides under development, but none has a major pharmaceutical sponsor.[48] According to UNAIDS, a first generation microbicide could be ready for distribution in as little as five to seven years.[49] A significant amount of investment and political will would be necessary to drive this. The Rockefeller Foundation estimates that roughly US$775 million would need to be invested in research and development to guarantee a successful product by the end of the decade. As of 2002, however, microbicide funding totaled only US$343 million.[50] I believe that a wide distribution of a successful microbicide could have huge impact on women and HIV/AIDS and must be put on the global AIDS agenda.

Conclusion

In this paper I discuss the many ways married women are especially vulnerable to HIV/AIDS. The circumstances involving migrant labor, lobola, and economic inequality lead to lowered agency and increased sexual risk behavior that in turn leave women highly susceptible to becoming the victims of infection. Effective policy must abandon current trends and put greater resources into female-controlled prevention.

In conclusion, I want to note that the subordination of women discussed in this paper by no means implies their weakness. The women I met in South Africa were some of the strongest women I have ever encountered. They bore and raised children, cared for the sick, endured tremendous suffering, held their families together, and still managed to be always singing. Women form the backbone of South African society and have a tremendous strength that – if channeled properly – could become a huge force for change.

Works Cited

Interviews and Participant Observation

Interview with Busisiwe Lekhuleni. Mthimba, South Africa, July 27, 2004.
Interview with Mumsy Sibeko. Mthimba, South Africa, August 12, 2004.
Interview with Jabulile Nxumalo, Masoyi office, August 17, 2004.
Discussion with Thanduxolo age 13, home of Florence Mbokazi, August 2004.

Books

Herdt, Gilbert, Ed. *Sexual Cultures and Migration in the Era of AIDS: Anthropological and Demographic Perspectives.* Oxford: Clarendon Press, 1997.
Lobola: It's implications for women's reproductive rights. Harare, Zimbabwe: Women and Law in Southern Africa Research Trust: Weaver Press [distributor], 2002.
UNAIDS / UNICEF / USAIDS Report: Children on the Brink 2002: A Joint Report on Orphan Estimates and Program Strategies. July 2002.
Von Kapff, Uli *Zulu: "People of Heaven."* South Africa: Holiday Africa Publications, 1997.

Journals and Articles

Albertyn, Catherine. "Contesting Democracy: HIV/AIDS and the Achievement of Gender Equality in South Africa." Feminist Studies 29, no.3, Fall 2003.
Caldwell, John C., John K. Anarfi, and Pat Caldwell. "Mobility, Migration, Sex, STDs, and AIDS: An Essay on Sub-Saharan Africa with Other Parallels." In Herdt, Gilbert, Ed. *Sexual Cultures and Migration in the Era of AIDS: Anthropological and Demographic Perspectives*. Oxford: Clarendon Press, 1997.
Campbell, Catherine. "Migrancy, Masculine Identities, and AIDS: The Psychosocial Context of HIV Transmission on the South African Gold Mines." In Kalipeni, Ezekiel, Susan Craddock, Joseph R. Oppong, and Jayati Ghosh, eds. *HIV and AIDS in Africa: Beyond Epidemiology.* Malden, MA: Blackwell Publishing, 2004.
Clark, Shelley. "Early Marriage and HIV Risks in Sub-Saharan Africa" Studies in Family Planning. Vol 35, No 3, Sept 2004.
Da Cruz, Claudia C. Bermudes Ribiero. "From Policy to Practice: the Anthropology of Condom Use." In: Kauffman, Kyle D. and David L. Lindauer, Eds. *AIDS and South Africa: The Social Expression of a Pandemic.* New York: Palgrave Macmillan Ltd, 2004.
Gregson, Simon, Constance A Nyamukapa, Geoffrey P. Garnett, Peter R. Mason, Tom Zhuwau, Michel Carael, Stephen K. Chandiwana, and Roy M. Anderson. "Sexual mixing patterns and sex differentials in teenage exposure to HIV infection in rural Zimbabwe." Lancet, Vol 359 (9,321) 2002.
Jochelson, Karen, Monyaola Mothibeli, and Jean-Patrick Leger. "Human Immunodeficiency Virus and Migrant Labor in South Africa." International Journal of Health Services, Vol 21, No 1, 1991.
Kalipeni Ezekiel, Susan Craddock, and Jayati Ghosh. "Mapping the AIDS Pandemic in Eastern and Southern Africa: A Critical Overview." In: Kalipeni, Ezekiel, Susan Craddock, Joseph R. Oppong, and Jayati Ghosh, Eds. *HIV and AIDS in Africa: Beyond Epidemiology.* Malden, MA: Blackwell Publishing, 2004.
Kelly, Robert, Ronald Gray, Nelson Sewankambo, David Serwadda, Fred Wabwire-Mangen, Tom Lutalo, and Maria Wawer. "Age differences in sexual partners and risk of HIV-1 infection in rural Uganda." Journal of Acquired Immune Deficiency Syndromes, Vol. 32, No. 4, 2003.
Klugman, Barbara. "Sexual Rights in Southern Africa: A Beijing Discourse or a Strategic Necessity?" Health and Human Rights, Vol. 4, No 2, 2000.

Lawson, Agathe Latre-Gato. "Women and AIDS in Africa: sociocultural dimensions of the HIV/AIDS epidemic." UNESCO, Blackwell Publishers, Oxford and MA, 1999.

Moleke, Percy. "The state of the labor market in contemporary South Africa." In:John Daniel, Adam Habib and Roger Southall, Eds. *State of the Nation: South Africa 2003-2004*. Cape Town, South Africa: Human Sciences Research Council, 2003.

Schoepf, Brook Grundfest. "AIDS in Africa: Structure, Agency, and Risk." In: Kalipeni, Ezekiel, Susan Craddock, Joseph R. Oppong, and Jayati Ghosh, eds. *HIV and AIDS in Africa: Beyond Epidemiology*. Malden, MA: Blackwell Publishing, 2004.

"South African Wimmin." The Economist, Vol. 341 Issue 1986, p79, 10/5/96.

Susser, Ida and Zena Stein. "Culture, Sexuality, and Women's Agency in the Prevention of HIV/AIDS in Southern Africa." In: Notes on: Kalipeni, Ezekiel, Susan Craddock, Joseph R. Oppong, and Jayati Ghosh, eds. *HIV and AIDS in Africa: Beyond Epidemiology*. Malden, MA: Blackwell Publishing, 2004.

Web Sources

Avert.org Adverting HIV & AIDS *http://www.avert.org/safricastats.htm*

Global Policy Network, Country Report: South Africa http://www.gpn.org/data/southafrica.html

Harter, Emily. "Women's Vulnerability to AIDS Key Element in UNAIDS Agenda." Washington DC. Allafrica.com/stories/200411160799.html

International Women's Health Coalition – Women and HIV/AIDS Fact Sheet *http://www.iwhc.org/resources/hivaidsfactsheet.cfm*

Loconte, Joseph. "The White House Initiative to Combat AIDS: Learning from Uganda." September 29, 2003. The Heritage Foundation. http://www.heritage.org/Research/Africa/BG1692.cfm

Masoyi HBC History of the Organization *www.masoyi.org*

Maharaj, Zarina. "Gender Inequality and the Economy: Empowering Women in the new South Africa." Keynote speech at Professional Women's League of KwaZuluNatal, August 9, 1999. www.africaaction.org/docs99/gen9908.htm

"Microbicides, Women and AIDS." UNAIDS The Global Coalition on Women and AIDS. www.unaids.org/html/pub/una-docs/GCWA_Microbicides_02Feb04_en_pdf/GCW...

UNAIDS 2004 Report section on "The Impact of AIDS on People and Societies." http://www.fightglobalaids.org/files/phatfile/New%20AIDS%20Statistics%20from%20UNAIDS.doc

UNAIDS 2004 Report on "Women and AIDS." http://www.unaids.org/wad2004/EPIupdate2004_html_en/Epi04_04_en.htm#P28_3962

Footnotes

[1] UNAIDS / UNICEF / USAIDS Report: Children on the Brink 2002: A Joint Report on Orphan Estimates and Program Strategies. July 2002.

[2] Avert.org Adverting HIV & AIDS http://www.avert.org/safricastats.htm

[3] UNAIDS 2004 Report section on "The Impact of AIDS on People and Societies." http://www.fightglobalaids.org/files/phatfile/New%20AIDS%20Statistics%20from%20UNAIDS.doc

[4] International Women's Health Coalition – Women and HIV/AIDS Fact Sheet http://www.iwhc.org/resources/hivaidsfactsheet.cfm

[5] Albertyn, Catherine. "Contesting Democracy: HIV/AIDS and the Achievement of Gender Equality in South Africa." Feminist Studies 29, no.3, Fall 2003, 597. (Women are said to be two times more likely than men to contract HIV from a single act of unprotected sex.)

[6] Lawson, Agathe Latre-Gato. "Women and AIDS in Africa: sociocultural dimensions of the HIV/AIDS epidemic." UNESCO, Blackwell Publishers, Oxford and MA, 1999, 393.

[7] This set of arguments is not exclusive to South Africa. South Africa as a nation shares many cultural and historical commonalities with neighboring countries, and the majority of my findings apply to regions throughout southern Africa. In this paper I look at some factors specific to South Africa as well as many broader topics that apply to married women across the region.

[8] For a review of past research mentioning married women and HIV see: Clark, Shelley. "Early Marriage and HIV Risks in Sub-Saharan Africa" Studies in Family Planning. Vol 35, No 3, Sept 2004, 150.

[9] Masoyi HBC History of the Organization www.masoyi.org

[10] Herdt, Gilbert, Ed. *Sexual Cultures and Migration in the Era of AIDS: Anthropological and Demographic Perspectives*. Oxford: Clarendon Press, 1997, 48.

[11] Caldwell, John C., John K. Anarfi, and Pat Caldwell. "Mobility, Migration, Sex, STDs, and AIDS: An Essay on Sub-Saharan Africa with Other Parallels." In Herdt, Gilbert, Ed. *Sexual Cultures and Migration in the Era of AIDS: Anthropological and Demographic Perspectives*. Oxford: Clarendon Press, 1997, 44.

[12] Campbell, Catherine. "Migrancy, Masculine Identities, and AIDS: The Psychosocial Context of HIV Transmission on the South African Gold Mines." In Kalipeni, Ezekiel, Susan Craddock, Joseph R. Oppong, and Jayati Ghosh, eds. *HIV and AIDS in Africa: Beyond Epidemiology*. Malden, MA: Blackwell Publishing, 2004, 144.

[13] Jochelson, Karen, Monyaola Mothibeli, and Jean-Patrick Leger. "Human Immunodeficiency Virus and Migrant Labor in South Africa." International Journal of Health Services, Vol 21, No 1, 1991, 164.

[14] Da Cruz, Claudia C. Bermudes Ribiero. "From Policy to Practice: the Anthropology of Condom Use." In: Kauffman, Kyle D. and David L. Lindauer, Eds. *AIDS and South Africa: The Social Expression of a Pandemic*. New York: Palgrave Macmillan Ltd, 2004, 153.

[15] Ibid, 152.

[16] Jochelson et al., op. cit, 165.

[17] Caldwell et al., op. cit, 45.

[18] Kalipeni Ezekiel, Susan Craddock, and Jayati Ghosh. "Mapping the AIDS Pandemic in Eastern and Southern Africa: A Critical Overview." In: Kalipeni, Ezekiel, Susan Craddock, Joseph R. Oppong, and Jayati Ghosh, Eds. *HIV and AIDS in Africa: Beyond Epidemiology*. Malden, MA: Blackwell Publishing, 2004, 60.

[19] Caldwell et al., op. cit, 48. My emphasis.

[20] *Lobola: It's implications for women's reproductive rights*. Harare, Zimbabwe: Women and Law in Southern Africa Research Trust: Weaver Press [distributor], 2002.

[21] Ibid, 12.

[22] Ibid, 28.

[23] "South African Wimmin." The Economist, Vol. 341 Issue 1986, p79, 10/5/96.

[24] Von Kapff, Uli *Zulu: "People of Heaven."* South Africa: Holiday Africa Publications, 1997, 37.

[25] Discussion with Thanduxolo, home of Florence, August 2004.

[26] Women and law in Southern Africa Research Trust, op. cit, 25.

[27] Klugman, Barbara. "Sexual Rights in Southern Africa: A Beijing Discourse or a Strategic Necessity?" Health and Human Rights, Vol. 4, No 2, 2000.

[28] Women and law in Southern Africa Research Trust, op. cit, 40.

[29] Da Cruz, op. cit, 157.

[30] Ibid, 139.

[31] UNAIDS 2004 Report on "Women and AIDS." http://www.unaids.org/wad2004/EPIupdate2004_html_en/Epi04_04_en.htm#P28_3962

[32] Clark, op. cit, 154.

[33] Lawson, op. cit, 394.

[34] Schoepf, Brook Grundfest. "AIDS in Africa: Structure, Agency, and Risk." In: Kalipeni, Ezekiel, Susan Craddock, Joseph R. Oppong, and Jayati Ghosh, eds. *HIV and AIDS in Africa: Beyond Epidemiology*. Malden, MA: Blackwell Publishing, 2004, 127.

[35] Gregson, Simon, Constance A Nyamukapa, Geoffrey P. Garnett, Peter R. Mason, Tom Zhuwau, Michel Carael, Stephen K. Chandiwana, and Roy M. Anderson. "Sexual mixing patterns and sex differentials in teenage exposure to

HIV infection in rural Zimbabwe." Lancet, Vol 359 (9,321) 2002, 903.

Kelly, Robert, Ronald Gray, Nelson Sewankambo, David Serwadda, Fred Wabwire-Mangen, Tom Lutalo, and Maria Wawer. "Age differences in sexual partners and risk of HIV-1 infection in rural Uganda." Journal of Acquired Immune Deficiency Syndromes, Vol. 32, No. 4, 2003, 446-451.

[36] Von Kapff, op. cit, 37.

[37] Maharaj, Zarina. "Gender Inequality and the Economy: Empowering Women in the new South Africa." Keynote speech at Professional Women's League of KwaZuluNatal, August 9, 1999. www.africaaction.org/docs99/gen9908.htm

[38] Moleke, Percy. "The state of the labor market in contemporary South Africa." In:John Daniel, Adam Habib and Roger Southall, Eds. *State of the Nation: South Africa 2003-2004*. Cape Town, South Africa: Human Sciences Research Council, 2003.

[39] Albertyn, op. cit, 598.

[40] Kalipeni et al., op. cit, 65.

[41] Participant Observation, Masoyi area, August 2004.

[42] Loconte, Joseph. "The White House Initiative to Combat AIDS: Learning from Uganda." September 29, 2003. The Heritage Foundation. http://www.heritage.org/Research/Africa/BG1692.cfm

[43] Lawson, op. cit, 399.

[44] Ibid.

[45] Susser, Ida and Zena Stein. "Culture, Sexuality, and Women's Agency in the Prevention of HIV/AIDS in Southern Africa." In: Notes on: Kalipeni, Ezekiel, Susan Craddock, Joseph R. Oppong, and Jayati Ghosh, eds. *HIV and AIDS in Africa: Beyond Epidemiology*. Malden, MA: Blackwell Publishing, 2004, 142.

[46] "Microbicides, Women and AIDS." UNAIDS The Global Coalition on Women and AIDS.
www.unaids.org/html/pub/una-docs/GCWA_Microbicides_02Feb04_en_pdf/GCW...

[47] Lawson, op. cit, 399.

[48] "Microbicides, Women and AIDS," Op. cit.

[49] Ibid.

[50] Ibid.

Jenny Tolan

Jenny Tolan is a senior majoring in International Relations and minoring in Spanish. Her academic interests include African and Latin American Studies, human rights, and international development. In the summer of 2004 Jenny spent 8 weeks in South Africa volunteering with the Masoyi Home-Based Care Project and doing research on women and AIDS vulnerability. Since writing this essay, Jenny had an article published about her research in the San Francisco Chronicle and has presented her policy suggestions as a fellow of the Center on International Development, Democracy, and Health of Stanford's Roosevelt Institution. Jenny spent her junior fall studying abroad in Spain at the University of Sevilla. She would like to thank Professor Terry Karl for her enthusiasm, support, and invaluable contributions to the research.

The Henry Cowell Student Lectures in Music:
Its Conception, Creation, Contiuation and Connection with Community

Yu "Carol" Cao

On the crest of every delightful knoll and around every beckoning corner, one encounters a different one of Stanford's "faces," each hued with unique and subtle nuances and yet each a shade of the same underlying essence. The whir of spinning bicycle wheels resonates against sandstone blocks of Spanish arches and slender trunks of palms marking the main quad. Light flows down sweeping glass curves and bounces off sleek steel railings of the Clark Center, dances through whispering leaves of trees in the Papua New Guinea Sculpture Garden, and directs flickering shadow screenplays on the ground below. Meanwhile, intellectual life boils within a flask as a reaction matures, floats on a powerful interpretation of a line of Shakespeare, and sings from a silent, still instrument suddenly infused with passion and vitality as skillful fingers depress and release black and white keys.

Focusing on this last image, one sees that it not only represents Stanford's appreciation for and cultivation of talent and potential nor only a striving for purity and excellence in musical performance. Also, the image depicts the highest form of originality and diversity. Different performers inevitably possess varying views on how to interpret or understand a certain piece of music; thus, although each musician may wield mastery over piano technique, each one's performance will possess his or her own individual style. Even when one performer's repeated portrayal of a particular piece is observed, subtleties of expression and execution manifest themselves. Creativity—of approach, of interpretation—contributes to this room for difference, transforming a chance mistake into an opportunity for spontaneous improvisation. Additionally, the musical selection itself is a conglomeration of diverse components, ranging from different notes and motifs to flexibility in execution, to the evocation of various moods in the listener, to the composer's intention, and to the background culture of the work.

Senior Jeffrey Treviño, a music major, expanded and applied the ideas delineated above to the diversity of musical traditions that he has observed on campus. In the course of exploring twentieth century piano music, he encountered composer Henry Cowell (1847-1965). Born in Menlo Park, Cowell grew up in the local area around Stanford and had many unusual and unconventional musical experiences throughout his life. Introduced to Iowa folk tunes and Irish songs and dances by his family, Cowell was also exposed to Gregorian chant and Cantonese opera in San Francisco and began violin studies at an early age. He never received formal piano training. However, in the spirit of experimentation, he was able to "generate sounds from [the instrument] that had never been intended by piano manufacturers."[1] Later, Cowell toured Europe, studied Oriental music in Berlin, and researched the musical traditions of Southern India, Japan, and Iran. He founded the New Music Quarterly (1927) to publish and disseminate the works of North and South American composers and the Pan American Association of Composers (1928) to promote and celebrate contemporary compositions.

Inspired by what he had learned about Cowell, Treviño began to examine his own environment at Stanford. He realized that Cowell's life work in musical innovation, his collection and appropriation of eclectic multicultural and pluralist components into original composition, and later musical review, criticism, and pedagogy paralleled the Stanford community's own inherently diverse musical backgrounds and identities. Yet, Treviño felt that something was lacking: "Although we [at Stanford] have a multitude of performance groups of various backgrounds—there isn't really a way for people to discuss diverse musical traditions, approaches, experiences, ideas in an academic context," he says.

This led him to formulate the idea for the Henry Cowell Lectures. Obtaining faculty support and funding through the music department, Treviño launched the project, selected Stanford undergraduate student presenters, and organized and executed the three-evening event (February 4-6, 2005). With free and open admission to the public, the "series hopes to present areas of music often left unexplored in academic

Jeffery Trevino plays the Tuba (Stanford News Service).

arenas and seeks to provide an engaging balance of both lecture and performance," Treviño writes on the Cowell lecture series website. Envisioning hands-on, interactive, interesting presentations in which the speaker performs actual demonstrations of the research methods, processes, or ideas that he is discussing, Treviño and a faculty committee evaluated submitted proposals and selected six for inclusion in the 2005 lectures. Topics covered musical traditions and innovations in composition, interpretation, and performance from the Renaissance to the present day, from the United States and Mexico to Germany and Japan.

Minna Chen, a Junior majoring in music and biological sciences, worked with Senior David Nuñez to present "Modern Pastourelles: Perspectives Eastern and Western, Male and Female." Originally an assignment for a seminar (Music 140: Studies in Medieval Music: Robin and Marion in Song and Story) taught by two-years' visiting professor and musicologist, Adam Gilbert, Chen and Nuñez had independently decided to examine and discuss parallels between modern expressive media and the standard pastourelle, a medieval poetic genre[2]. While Nuñez looked at lyrics of works by contemporary musical artists, Eminem and Sublime, Chen explored ways in which to interpret different views of male-female relationships through the lens of the pastourelle, focusing specifically on rape and female subjugation as portrayed in Raise the Red Lantern, a modern Chinese film that incorporates the same general features (i.e. character development, plot

elements, musical structures) as the standard pastourelle. Through her research, Chen realized that the flexibility of the pastourelle allows for its adaptation to any culture and time. She had also come to see that "the reason the pastourelle is so powerful, even today, is that it serves as a commentary on cultural realities and attitudes toward gender relationships."

Senior Mahan Esfahani, majoring in music with a concentration in musicology, studied the development of the 15th century French chanson, approaching the subject from textual, musical, and poetic vantage points, similarly in a seminar taught by Professor Adam Gilbert. During this time, Esfahani joined the new performance group Ciaramella as an organist after his introduction to it by Professor Gilbert, one of the world's foremost performers on Renaissance wind instruments (recorders, shawms, bagpipes). Thus, he had an opportunity to work with selections of 15th century keyboard music, collected in the Buxheim Organbook (compiled in southern Germany, ca. 1455-70). Focusing on one piece in this collection—Guillame Dufay's famous chanson "Se la face ay pale"—Esfahani undertook a quarter-long study to analyze the compositional technique involved in the translation—or "intabulation"—of the work, originally based on a vocal model, into a form suitable for keyboard performance. "Musical scholarship and performance are not and ought not be independent of each other," Esfahani states, "A big part of my research is indeed [motivated by] what I feel—I want to contribute to the crusade against the gap I see between scholarship and performance perpetuated by the insecurities of musicologists and performers alike." These beliefs compel Esfahani to convey his musical research and interests to community audiences through performance. Adept at a wide variety of keyboard instruments, Esfahani finds gratification in helping to dispel a commonly upheld misconception that early music is not good music. "A little background information goes a long way, and audiences are truly convinced," he comments, "What's amazing is showing people what makes [a particular musical genre] compelling on its own terms—like illuminating a work of art by its own light."

Junior Yoko Okano, a double major in music and psychology, is the recipient of an Undergraduate Research Opportunities (URO) grant. She proposed to study the historical roots and growth of Japanese and North American kumidaiko (ensemble drumming) through the influence of Daihachi Oguchi sensei[3], an innovator of the concept in the early 1950s, and his group Osuwa Daiko. With the support of her faculty advisor, Stephen Sano, an associate professor in music, and Linda Uyechi, a lecturer in taiko[4], Okano traveled to Japan last summer to interview

Oguchi sensei and several prominent figures in the taiko community there. Additionally, Okano had the opportunity to attend several taiko workshops led by Oguchi sensei himself and thus see and experience firsthand his charisma as a teacher and artist. From her research, Okano plans to write a comprehensive and in-depth history of the beginnings of kumidaiko in Japan, focusing on Oguchi sensei's impact and to produce an annotated bibliography of taiko references in both English and Japanese that may become a valuable resource to the general taiko community, the Japanese American National Museum in Los Angeles, the Kodo Arts Sphere America (KASA), a non-profit organization that seeks to encourage the study and understanding of kumidaiko (http://www.kodo.or.jp/kasa/frame.html), and the Introductory Sophomore Seminar on Perspectives of North American Taiko at Stanford (Music 17Q). Given her active participation in Stanford taiko and her artistic and academic interests in its performance and psychological and social contexts, Okano hopes that her research will contribute to the currently available body of English literature on the young and developing but now international art form of kumidaiko.

With combined interests in film, television, and popular (i.e. pre-existing) music and song lyrics, Yogesh Raut, a Senior majoring in psychology, researched the re-appropriation and incorporation of song lyrics into popular culture. "My research grew out of noticing how popular song lyrics can express complex emotions in ways that other art forms cannot, and how coupling this effect with visual images and a story can contribute to the creation of even stronger and more complex emotions," he says. Initially noting some of the more apparent uses of songs in film including their roles in orienting the viewer as to the film's setting, in adding historical, social, or local color, and in introducing the attractive element of performance to a film, Raut went on to observe perhaps less obvious uses of songs in film. For instance, Elton John's "Amoreena" helps to establish and clarify mood in Sidney Lumet's film Dog Day Afternoon, while Simon and Garfunkel's "Sounds of Silence" evokes, clarifies, and explicates themes presented in Mike Nichols' The Graduate. Raut demonstrates how repetition of James Taylor's "Fire and Rain" at different points of Sidney Lumet's film Running On Empty creates a motif that further underscores the themes involved and how films such as Peter Hunt's On Her Majesty's Secret Service and Michael Moore's Bowling for Columbine use the songs, "We Have All the Time in the World" and "What a Wonderful World," respectively, by Louis Armstrong to produce an ironic effect. Through detailed analysis of the full lyrics of songs including "You Might Be My Lucky Star," "Get Outta My Dreams, Get Into My

Taiko workshop given by Osuwa Daiko (Oguchi sensei's group) at the Festival at the 2004 Okava Taiko Festival in Japan.

Car," and Belinda Carlisle's "Heaven Is a Place on Earth" on the soundtrack of the film Safe, Raut discovered that including popular songs in films has the effect of drawing inter-textual connections and thus, expanding the film as a text. Additionally, Raut looks at how songs can create a sense of unity, family, and community through group performance and group enjoyment or appreciation, in the film's characters, cast, producers, and audience—for example, The Foundations' "Build Me Up, Buttercup" in Peter and Bobby Farrelly's There's Something About Mary. Ultimately, Raut hopes that his research will "shine a light on an under-studied and under-appreciated aspect of film and music scholarship and help to bring together the worlds of film and music."

Daniel Herrera's presentation entitled "Jalisco, Veracruz, and La Huasteca: Musical History of the Three Mexican Son Traditions" provided his audience with a "tour of Mexico's rich mariachi, jarocho, and huasteco traditions,with L.A.'s multi-genre ensemble Mariachi Nuevo Cuicatlán."[5] A political science major and double minor in music and CASA (Cultural and Social Anthropology), Herrera examined the three primary influences of Mexican son indigenous, European, and African musics, showing his audience how each one of these influences was emphasized in a different geographical region, with "indigenous traits appearing most often in the rural Huasteca coast of Mexico, jarocho traditions holding on to African rhythms on the eastern Veracruz coast, and European aesthetics characterizing mariachi music in the western coastal state of Jalisco." Herrera demonstrated

how the regional isolation and environment allowed for each son tradition to evolve independently of each other and yet still retain the basic son meter of the sesquialtera, a meter alternating between 3/4 and 6/8.

Although this was the incipient year for the lectures, Treviño feels that the project has been a success. "I can say that I feel like I've accomplished what I set out to do. The range of presentation topics spanned a wide variety of times, cultures, and musical styles. Furthermore, the presenters have communicated their ideas with the assistance of performance ensembles—the audience itself in the case of Carolyn Chen's[6] lecture on free improvisation. These are my kind of lectures," he says. Treviño has high hopes

that the lectures will become a repeated annual event in the future, providing the Stanford and surrounding community with new and diverse perspectives on music and musical research and an open forum for discussion of such topics. Daniel Herrera, one of the lecturers this year and featured above, will be organizing this event in 2006 as a Senior.

The author would like to thank Jeffrey Treviño and each of the 2005 student lecturers featured in this article (Minna Chen, Mahan Esfahani, Yoko Okano, Yogesh Raut, Daniel Herrera, Carolyn Chen) for their large contributions to this article.

Footnotes

[1] The Henry Cowell Lectures website is located at http://music.stanford.edu/Events/cowellLects/.
[2] The pastourelle is a 14th century poetic genre, usually written in French. General defining features are its presentation in pastoral literary mode, a cast including a man and a woman, a plot involving discovery and attempted seduction, structure with both narrative and dialogue, a masculine first-person point of view, and a possible refrain.
[3] Sensei is a term of respect for teachers in Japanese culture.
[4] Taiko means drum in Japanese. Kumidaiko literally means to play taiko together with others.
[5] Quote taken from the Cowell lectures website
[6] Carolyn Chen's presentation included discussion of music by Pauline Oliveros, Cornelius Cardew, Christian Wolff, and her own audience—inviting the audience members to create their own music by means of instruments such as kazoos and slide-whistles; this interaction contributed to a fun and innovative lecture.

Special Features

A New Approach to HIV Prevention in Uganda

Anant Ramesh Patel

An estimated 25 million adults and children were living with HIV in sub-Saharan Africa at the end of 2003, and an estimated twelve million children have been orphaned by AIDS. In 2003, 2.2 million people died from AIDS. According to the Uganda AIDS Commission, 78,000 of these deaths were in Uganda and about 940,000 children in this country have been orphaned[1]. There is no doubt that H.I.V is a serious problem that needs to be addressed. In the past, Ugandan community groups and NGOs have attempted to educate the public about HIV through basic informational campaigns, hoping that an increase in awareness will bring about behavioral change. However, it is becoming increasingly clear that handing out pamphlets showing statistics on H.I.V and distributing condoms is simply not effective in encouraging sustained

behavioral change.

This past summer, Stanford Undergraduate Leila Ehsani went to Uganda to learn about novel approaches to HIV education that propose to effect change in regards to adolescent sexual practices. Her research write-up entitled "HIV Edutainment in Uganda: Communicating a Behavioral Change HIV Message to Youth" conducts a within-and-between case study analysis of three modes of edutainment in Kampala, Uganda— print, radio, and live drama—in order to understand how these programs are attempting to encourage successful behavioral change in regards to HIV among youth. The following represents an interview with Leila about her research and experience in Uganda.

What exactly is edutainment? What is the entertainment component of the edutainment programs that community groups and NGOs use?

Edutainment describes the combination of education and entertainment in behavioral change programs. In Uganda, several community groups and NGOs are trying to incorporate ideas of abstinence and safe-sex into comic books, radio broadcasts, and live dramas that are performed in local schools by local residents. The performances consist of live testimonies of people living with H.I.V as well as skits that depict the challenges experienced by Ugandan youth regarding HIV. This approach seeks to be highly interactive with youth.

Why edutainment?

People are bored of being lectured to about H.I.V. You can't just stand up and talk about why HIV is bad and why it needs to be avoided because youth are already aware of this. Thus traditional methods might increase awareness, but are most likely less effective in encouraging behavioral change. Edutainment is designed to be a communication tool that bridges this gap between awareness and behavioral change.

What does edutainment need to address in order to be successful?

One aspect that the involved community groups and NGOs emphasize is that it is extremely important to not only reach out to youth, but also to those who influence them most (parents, teachers, etc). These are key people that need to be incorporated into the process. By involving
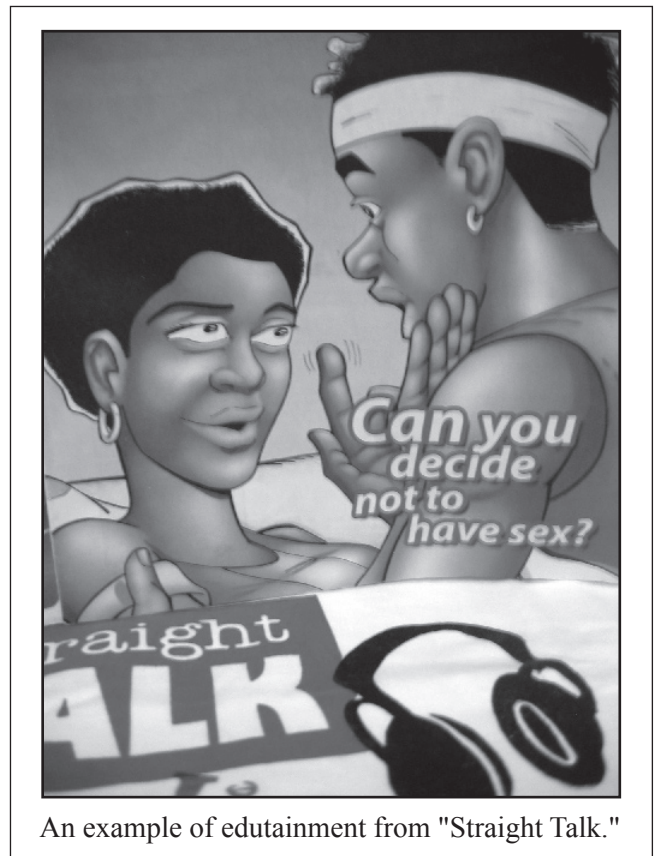
An example of edutainment from "Straight Talk."

people youth interact with on a fundamental and daily basis, it is believed that the involved youth are more likely to internalize the message. Remember, it is the parents and teachers that meet with youth for the majority of the day and influence them most in their decision-making. A suggested way of making this involvement possible is by having information pieces in the national newspaper that the parents or teachers can share with youth. Another initiative that has been explored is to encourage teachers to hold regular classes on HIV/AIDS. Finally, it is absolutely essential to incorporate youth into the entire process, from planning to evaluation. If they are made to feel as though they own the process, edutainment programs are likely to be much more influential on their lives and decisions.

What conflicts/barriers are there to edutainment's success?

One interesting challenge that I noticed was the conflict between science and religion regarding HIV/AIDS in Uganda. Scientific evidence says one thing, and religion advises on another. Some religious communities, for example, say things like "condoms will give you H.I.V" and "if you join our community, you'll be immune to H.I.V." Such statements make it immensely harder to make educational practices effective, especially since religion plays such a seminal role in the majority of the population's lives. If HIV education efforts in Uganda are

Leila and Ugandan youth.

to be effective, science and religion must be in harmony.

How has edutainment succeeded so far?

It is difficult to say because the programs are relatively new. Moreover, there are limited resources, so setting up evaluating mechanisms is difficult. The only certain thing is that traditional approaches to the HIV problem are no longer good enough. From my experience in Uganda, it seems that even if edutainment programs succeed in conveying a behavioral change HIV message to youth, they will have a limited effect on behavior if they do not seek to transform society at large. If girls are expected to act in ways that render them vulnerable to HIV, if poverty continues to constrain the choices that young people make, and if a supportive home and school environment is lacking, no matter how effective HIV edutainment might be as a communication tool, its efforts will be in vain.

Leila Enhansi is a junior majoring in Human Biology. Originally from Kenya, she hopes to return to Africa to be of service in the field of public health and development.

Special Features

A Blessing

Vanya Choumanova

In the spring quarter of 2003-04 academic year, I participated in Stanford's Overseas Studies Program in Santiago, Chile, where I had the incredible opportunity to immerse myself in a new culture. My explorations of the Chilean health care system through volunteering in a publicly funded orphanage and visiting patients in a cancer ward were two of the most meaningful experiences of my undergraduate career at Stanford University. These experiences motivated me to return to Chile and to carry out a research project through which I yearned to help patients in their struggle against breast cancer.

I obtained a URO Major Grant and went back to Chile in December 2004. During my four-week long stay in Santiago, I was able to build a very close relationship with both patients and doctors at the Breast Cancer Division at Hospital Barros Luco. By conducting interviews and using three psychometric scales (Mini-MAC, RCOPE, and FACT-Sp), I investigated how female Chilean cancer patients cope with breast cancer and how religion is used as a coping mechanism. The preliminary findings of the study demonstrate that religious coping is an essential coping mechanism for the majority of the informants. As one study participant reported, "Los únicos recursos que tenemos aquí son nuestras creencias y nuestra fe…La religión y la fe espiritual me ayudan a sobrellevar mi enfermedad." (The only resources we have here are our beliefs and our faith. Religion and spiritual faith help me bear my illness.)

I was struck by the differences between the resources available to patients in the United States and Chile and how those differences affect patients' coping mechanisms. In the Chilean public medical institution, patients did not rely on scientific technology or medication to be cured. Such resources were often absent or beyond the financial reach of patients. Instead, breast cancer patients at Hospital Barros Luco prayed, made offerings to Saints, and searched for spiritual help from their pastors as they hoped to recuperate. None of the study participants blamed God for their diagnosis with breast cancer. In fact, many of the patients even referred to cancer as a blessing, God's way of allowing them to enjoy and value what is important in life.

Despite my enthusiasm and dedication to this research project, there were hurdles along the way, as should be expected for any domestic or overseas research projects. First, Spanish was not my native language. I understood the majority of what informants were telling me, but some valuable information was certainly lost in the interviews. Second, the quantitative portion of the study that I had carefully designed with the help of my research advisors at Stanford was foreign and difficult for some of the study participants. Finally, it was during winter break that I was in Chile; events at the hospital were not running as usual because of the Christmas holiday season.

Creativity helped me overcome these hurdles. I managed to complete the data collection. The investigation

Breast cancer patients participate in occupational therapy and sell their arts and crafts at annual exhibitions. The money they raise is used to buy medications for cancer patients who cannot afford drugs.

Two teams of volunteers (Damas de Verde and Damas de Azul) offer emotional support in Hospital Barros Luco by sharing their personal experiences with breast cancer.

allowed me to satisfy the curiosity stemming from my personal and academic interests in the topic. However, most importantly, I developed a passion for the people. I cared for the patients as I care for my own family and friends. It was insignificant that I was a foreigner and a stranger. In fact, that made it easier for a lot of patients to share with me their deepest fears about their illness and their frustrations with the public health system. They trusted me and looked up to me as an educated woman who had the potential to make a change. Some of the women thanked me for listening to them; they thought the interviews were therapeutic especially since it was difficult for many of them to speak about cancer with family members.

After speaking to more than 30 breast cancer patients and the medical staff, I knew that I must devote my career to cancer eradication. As soon as I got back to the States, I established a connection with American Cancer Society. I sought help by expressing to American Cancer Society that the needs of this community were unmet, and that thousands of patients needed financial support to buy medication. Although my research experiences were life-altering, my work in Chile is just the beginning. I have

listened to patients' complaints, frustrations, and fears; now I have to take it upon myself and do something with that information. I did not go to Chile to conduct research so that I could complete my Honors thesis. I went to Chile with two missions: to make a personal connection with the community and to use my skills and knowledge in ways that are useful to that community. My trip to Chile has been a journey of self-discovery. In a way, it is my blessing.

Vanya Choumanova is a senior majoring in Human Biology and minoring in Spanish Language. A native of Bulgaria, she moved to the United States at age fourteen but quickly learned both English and Spanish through her bilingual and ESL classes in NY. Vanya would like to thank Dr. Cheryl Koopman, Dr. Ronald Barrett, and Dr. Stan Wanat for their advice, mentorship, and encouragement through all stages of the research project. Vanya plans to become an oncologist, but she would first like to carry out a public service project and implement a health education program in Hospital Barros Luco during her time off before medical school.

Family Acceptance Project

Zaw Lin Hteik

Stanford Pride is an organization whose mission is to create and foster a diverse LGBTQQI (Lesbian, Gay, Bisexual, Transgender, Queer, Questioning, and Intersex)-affirming community of alumni, students, faculty and staff to guide Stanford to become a model institution on LGBTQQI issues and to support Stanford students. Every year, they award at least one Stanford Pride Fellowship to current Stanford students who are interested in spending a summer addressing LGBTQQI issues by working fulltime with an organization actively engaged in this area of work. In the summer of 2003, Adam Forest, a sophomore in 2002-2003 academic year, was awarded the Stanford Pride Fellowship. He used the fellowship to join the Family Acceptance Project being conducted by San Francisco State University. The following is an interview with Forest about his experience and thoughts on the project he was a part of in the summer of 2003.

Q: How would you briefly describe the research/project that you were a part of in summer 2003?

A: Growing up as a queer youth in this country is no cakewalk. Few are lucky enough to have parents that are not outright rejecting; those who do are sure to face ridicule and sometimes hatred from their peers and the larger community. And—though it be perhaps and obvious point—LGBT people do not reproduce: queer youth are almost always forced to seek their identity alone, without the help of supportive culture or family. It should come as no surprise, then, that queer youth have elevated rates of depression and dropout; suicide rates for queer youth are 4 times that of average youth, not counting those who never had the courage to come out before completing a suicide.

The Family Acceptance Project is the first research project to study queer youth and their families together. Aimed at increasing the resiliency of LGB youth, the project looks at what factors contribute to the psychological an physical health of LGB teens ages 13-18. It also seeks to discover what types of knowledge parents need in order to help them be more accepting and supportive of their queer adolescents. When finished, the project will provide tools to health care professionals and community service providers to assist LGB youth and their families in living healthier lives.

Q: What was your role in this project?

A: As a summer intern, I did a number of things for the project. Graphic design was my favorite part of the work that I did: in the end I designed a booklet that was printed as a recruitment tool

Adam and FAP (Family Acceptance Project) members.

to persuade LGB youth and their families to participate in the study. I also spent hours pouring through audiotapes of the interviews that were being conducted with LGB youth and at least two of their family members, listening for common themes and experiences. On top of this I did a variety of other smaller jobs, like interacting with the Board of Advisors and recruiting youth at the San Francisco Pride Parade.

Q: How has this experience shaped your Stanford education?

A: LGBT activism is something that is very important to me, personally and academically. As a straight white male, it is very easy to make heads turn when people learn about my involvement in LGBT work. My continued activism, from working at the LGBT-CRC on campus to researching with the Family Acceptance Project of SFSU, has enabled me to be educated about the realities of being queer in contemporary American culture, and has informed much of my study related to civil rights law and political philosophy.

Q: Has this experience changed any of your perspectives towards LGBT community and if yes, how?

A: Going over all those audiotapes, it was striking just how similar many of the LGBT youth's experiences were. In fact, they were nearly identical at times. The coming out experience, for example, was an incredibly intense moment for all of them, characterized by fear and uncertainty. This was especially the case when coming out to one's parents. On the other end, the experience of parents whose children came out to them was also very similar. Many of them brushed it off as a phase or something that they child would grow out of; this was often cited as a source of frustration and hurt for many LGB youth, who wanted only to be taken seriously by their parents.

Q: What was the highlight of this project experience?

A: The highlight for me was working with a staff of very dedicated, intelligent staff and faculty from SFSU. They are completely committed to the project and all shared an intense desire to reshape the experience of queer youth toward a more livable adolescence. It is uncommon, I think, that individuals possess the desire to positively affect a group of people that is not their own. The staff of the Family Acceptance Project stood to gain nothing personally from improving the lives of gay youth; in this sense it is substantially different from fighting for gay marriage or protection from discrimination. Knowing them was a blast, and a privilege.

More about the project Adam was involved in can be found at http://familyproject.sfsu.edu/index.htm

Adam is currently a senior at Stanford University majoring in Philosophy and is writing his honors thesis on sex change and health insurance in Ethics in Society. For the past two years, he was on staff at the LGBT.CRC (Lesbian, Gay, Bisexual and Transgender – Community Resource Center) at Stanford University. He is also currently a residential assistance at Theta Delta Chi on campus. Adam is a founder and the current Editor-in-Chief of Six Degrees, A Stanford Journal of Human Rights. He is planning to take a year off and move to Buenos Aires after he graduates. After his stay in Buenos Aires, Adam plans to attend law school in United States and study civil rights law so that one day he may be a part of rolling back the discriminatory legislation against LGBT people.

Special Features

Family Resource Desk:

Connecting Patients with Community Resources

Nancy Wang

Stanford undergraduates working at the Family Resource Desk (FRD) strive to address the non-medical but clearly health-related needs of Stanford Emergency Department (ED) patients and their families. Staffed and coordinated by Stanford undergraduates, the FRD serves the information needs of children and their families by connecting them with community resources and services.

The resource desk was the brainchild of Dr. Ewen

Wang, Assistant Professor of Surgery in the Division of Emergency Medicine at Stanford. "I am very excited about the entire concept of the FRD," said Dr. Wang. "As a physician, I am so busy with patient care, I am unable to give patients accurate information about programs, let alone solicit other worries, stresses or needs even though they directly affect their health. Within the FRD, our undergraduates provide direct service to individual patients, research social programs, as well as gain valuable exposure to the myriad of factors that affect health within the ED setting."

Many Stanford ED patients are from East Palo Alto, which, according to the US Census bureau, consists of over 70% minorities and has an average income less than $14,000 per year. According to Prasanna Ananth, a second year medical student at Stanford and founder of the FRD, the combination of demographic factors and barriers to obtaining important social services contributes to health care disparities that the FRD hopes to address.

Ananth and several Stanford undergraduates conducted needs assessment surveys during the summer of 2004 to determine the types of information that would be helpful to families in the ED. The assessment found that more than a third of the patients in the ED were at or below the Federal Poverty Line, currently at an annual income of $18,850 for a family of four. Ananth and students also found that 70% of the patients were ethnic minorities, and 61% had received high school education or less. These findings reinforced the observation that many of the families of pediatric patients seen in the ED are from disadvantaged backgrounds. The assessment also confirmed patient interest in access to resources in the ED itself. Informational needs expressed by patients included details of educational opportunities for children and adults, housing and health insurance. "It became apparent, following my analysis of the needs assessment data, that families felt disconnected from public resources and would appreciate a help desk such as our own to more effectively utilize the social services available to them," said Ananth. "This essentially validated our mission in initiating the desk."

The FRD was initiated as a part of Ananth's Valley Foundation Medical Scholars project in the Stanford School of Medicine, and initial logistics and organizing were done largely by Dr. Wang and Ananth. Wang and Ananth recruited undergraduates in April of 2004 to conduct preliminary research. Over 70 students expressed interest, and around 15 undergraduates formed study groups to examine the key topics of health care access, nutrition and food stamps, shelters, continuing education, and child care services. The Valley Foundation provided initial start-up funding for the organization, although the FRD is now registered as a volunteer student organization and will receive funding from the Associated Students of Stanford University (ASSU).

Since the opening of the desk in November of 2004, the number of volunteers at the desk has grown to 24 individuals who work 4-6 hours per week at the desk. Over 20 patients have been comprehensively served. Three undergraduate coordinators, selected annually, are also responsible for overseeing the day-to-day operations of the desk, meeting with staffers, publicizing the desk, and considering future directions of the program. Focusing on children with families, staffers introduce the program to patients in the ED, ask if patients need any specific assistance and present resources available. Patients are also referred to the desk by medical care teams who, according to the program coordinators, have been very responsive. Staffers follow up with patients to ensure that they were able to establish contacts or obtain the necessary services. Based on patient requests, volunteers also continuously update information and assess the need for new resource materials, which are then researched, compiled and incorporated into the FRD database.

Areas of current resources include low-cost community clinics such as Arbor and Pacific Free Clinics, health insurance, nutrition, adult education, obesity, child care, temporary housing, subsidized housing and child and crisis counseling. Sophomore Sherveen Salek, one of the coordinators, finds health insurance an area where the FRD can contribute greatly to a child's health and well-being. "Since essentially all children within the San Mateo and Santa Clara counties are eligible for health insurance, the
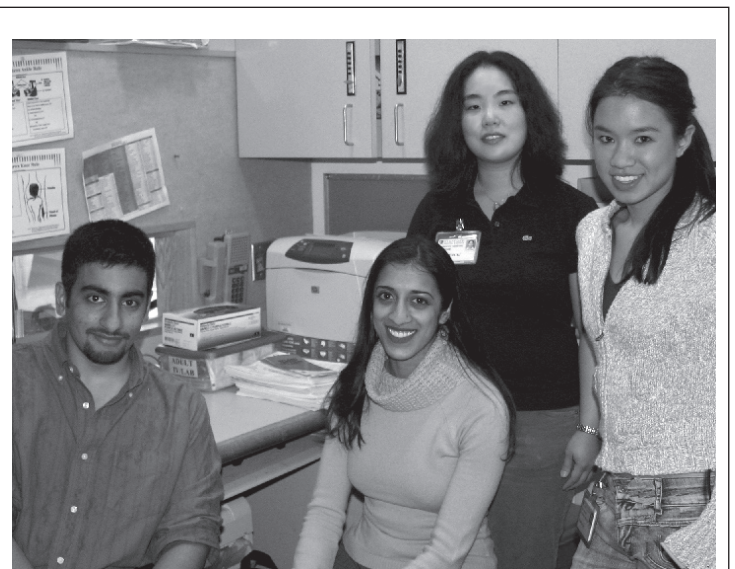
Sherveen Salek, Prasanna Ananth, Mari Suzuki and Jessica Li at the Family Resource Desk in the Stanford Emergency Department.

FRD plays a crucial role in informing patients of the options available to them," he said.

The FRD hopes to continue to increase the number of staffers through quarterly recruitments, and this year's coordinators are looking into expanding the program. Sophomore Jessica Li, another coordinator, said, "Many undergrads are interested in our program because not only does it offer them a valuable learning experience – whether they are bound to medical school, law school, or even to the business sector – it also provides them with the opportunity to give back to the community." Students who are interested in volunteering or learning more about the program should contact coordinators Jessica Li (jessli@stanford.edu), Sherveen Salek (Sherveen@stanford.edu) or Mari Suzuki (mari.suzuki@stanford.edu). Dr. Wang and the undergraduate coordinators will also be teaching a student-initiated course in the spring entitled "Community Health, Contemporary Issues in Health Care." "Through the student-initiated course, we hope to foster more awareness of the socioeconomic factors profoundly affecting the health of many Emergency Room patients," said Suzuki.

As the FRD continues to provide and expand services to the underserved community at the Stanford ED, it demonstrates that undergraduate research can have a large social impact. The program also serves as a reminder that research is most frequently an on-going process that depends on dedicated individuals working to further knowledge and thereby enable positive change.